1984

Mixed model procedure for the unification of within herd evaluations through national beef sire evaluation

Doyle Edward Wilson
Iowa State University

Wilson, Doyle Edward

MIXED MODEL PROCEDURE FOR THE UNIFICATION OF WITHIN HERD EVALUATIONS THROUGH NATIONAL BEEF SIRE EVALUATION

Iowa State University PH.D. 1984

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark __√__.

1. Glossy photographs or pages _____
2. Colored illustrations, paper or print _____
3. Photographs with dark background _____
4. Illustrations are poor copy _____
5. Pages with black marks, not original copy _____
6. Print shows through as there is text on both sides of page _____
7. Indistinct, broken or small print on several pages __✓__
8. Print exceeds margin requirements _____
9. Tightly bound copy with print lost in spine _____
10. Computer printout pages with indistinct print _____
11. Page(s) _____ lacking when material received, and not available from school or author.
12. Page(s) _____ seem to be missing in numbering only as text follows.
13. Two pages numbered _____. Text follows.
14. Curling and wrinkled pages _____
15. Other _____

Mixed model procedure for the unification of within herd evaluations through national beef sire evaluation

by

Doyle Edward Wilson

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Department: Animal Science
Major: Animal Breeding

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1984

TABLE OF CONTENTS

SYMBOLS AND ABBREVIATIONS

A	Wright's (1922) numerator relationship matrix
BLUE	Best linear unbiased estimate
BLUP	Best linear unbiased prediction
CPU	Central processing unit
E	Expected value
EPD	Expected progeny difference
I	Identity matrix
h^2	Heritability
NID	Normally and independently distributed
NSE	National sire evaluation
$r_{G_D G_M}$	Genetic correlation between direct and maternal traits
K	Computer memory equal to 1,024 bytes
PEV	Prediction error variance
PL/I	Programming language one
WWT	Weaning weight
Var	Variance
$\otimes$	Direct matrix product
Σ	Summation
σ_i^2	Variance for effect i
σ_{ij}	Covariance between effects i and j

INTRODUCTION

Artificial insemination (AI) and high speed computers have enabled beef cattle seedstock and commercial producers to witness a major revolution in live animal evaluation. Adoption of AI as a means of incorporating new germ plasm into a breeding herd has made possible the development of evaluations that today are national in scope. AI sires allow direct comparisons to be made between thousands of sires that otherwise would not be possible. The high speed computer also makes possible the analysis of hundreds of thousands of progeny performance records collected from across the continent in a national sire evaluation.

Other technology has progressed concurrently with the use of AI and computers. Animal breeding geneticists and statisticians have developed evaluation methodologies that account for the biological nature of the germ plasm pool. New evaluation methodologies also account for environmental factors that influence phenotypic expression of economically important beef cattle traits.

Live animal evaluation has not, however, reached a state of perfection even with the advancements made in the last decade. Researchers continue to seek improvements in evaluation methodology and elimination of evaluation limitations. Current areas associated with beef sire evaluation research are: multiple trait evaluation, interaction among fixed and random effects, variance component estimation, and adjustment factors for herd genetic merit in intraherd prediction.

A new horizon in the improvement of genetic evaluation procedures and methodology is the development of methods to unify the national

beef sire evaluations with another aspect of seedstock evaluation, within herd evaluations of dams and young animals. A primary reason for unifying within herd evaluations and national sire evaluations (NSE) is to obtain genetic evaluations of unproven yearling bulls, bulls with only their own performance and no progeny records.

NSE is a progeny test of AI sires that leads to high accuracy in a sire's evaluation. However, improvements to accuracy are offset by the inherent increase in generation interval. Only sires with moderately high accuracy get national visibility in sire summary reports. Young yearling bulls with no progeny have no opportunity for across herd comparisons and have no opportunity for national visibility. In addition, young AI sires (two and three years old) have little opportunity of obtaining national visibility in NSEs due to their low accuracy values.

The AI sires with progeny in different contemporary groups provide the common base for fair comparisons among sires. These sires establish the covariance ties between bulls in the NSE. The only way to tie together young yearling bulls without progeny is with the active AI sires that have progeny across herds and contemporary groups. Therefore, genetic evaluation of sires from the NSE for these active sires becomes the reference point from which young yearling bulls in individual herds can be fairly compared. Given the genetic breeding value of a yearling bull's sire, determined from an NSE and the dam's breeding value from a within herd evaluation tied to the NSE, a yearling bull's breeding value can be determined and directly compared to the breeding value of a yearling bull from another herd.

Unification of national beef sire evaluation and within herd evaluations has the potential of improving current evaluation procedures. The purpose of this thesis is to present and demonstrate with an actual set of weaning weight performance records one approach to the unification of national beef sire evaluations and within herd evaluations. The approach modifies current sire evaluation procedures by including dam effects as well as sire effects in the evaluation model. The approach uses best linear unbiased prediction (BLUP) procedures for both the sire and within herd evaluations. A selected set of 20 purebred Angus herds is used to validate the evaluation procedures and to determine computer requirements for an operational system. The review of literature that follows gives a brief historical perspective of the development of current evaluation procedures. In addition, the review points out some deficiencies of current evaluation procedures and alternative approaches to eliminating these deficiencies.

REVIEW OF LITERATURE

The methodology base from which current beef sire evaluations are derived was developed first in support of dairy sire evaluations. Dairy breeders began to make significant genetic advances in the late 1950s and the 1960s through use of herdmate comparisons. Gains in genetic improvement were great enough to invalidate the underlying assumptions in the procedures of herdmate comparisons. This in turn rendered the evaluations less useful (Freeman, 1980). One assumption that became invalid was the assumptions that all sires came from a common genetic population. When genetic advances occurred, distinct genetic sub-populations began to emerge, and the herdmate comparisons could not account for the genetic trend.

Henderson (1966) suggested that dairy bulls undergoing evaluation should be divided into groups for evaluation. A bull's evaluation would be the sum of the estimate of his group and a selection index type of evaluation of his deviation from the group mean. Henderson referred to this as mixed model selection following the analogy of mixed model analysis of variance.

The use of mixed models actually predated this period of changing dairy sire evaluation procedures to account for genetic trend. Henderson et al. (1959) proved that solutions for fixed effects in the linear mixed model gave solutions identical to those obtained by generalized lease squares. Henderson (1963) also proved that solutions for the random effects from the mixed model are best linear unbiased predictions.

Refinements to the mixed model formulation, determination of model properties, and applications to sire evaluation can be found in a series of Henderson's research articles (Henderson, 1973, 1974, 1975a,b,c,d,e).

In the mid-1970s, beef cattle breeding researchers began to explore the benefits to be gained in using Henderson's mixed model procedures to evaluate beef sires. Nielson (1974) examined seven alternative analysis procedures to estimate national sire expected progeny difference (EPD) values. Although differences in sire EPDs between analysis procedures were small, the mixed model evaluation procedure was the method of choice due to its theoretical properties. Another application of mixed model procedures was an evaluation of sires from different breeds (Schaeffer and Wilton, 1975). Schaeffer and Wilton chose the mixed model approach because the model was able to account for many of the factors influencing beef growth traits.

Pollak et al. (1977) looked at alternative mixed model formulations as the means of providing within herd sire evaluations to beef cattle breeders. They studied mixed model procedures for within herd sire evaluations because using progeny averages (ratios derived from deviations from contemporary group averages included) did not effectively compare sires with varying numbers of progeny. In addition, use of progeny averages did not effectively rank animals with records made at different time periods. Pollak et al. (1977) also pointed out that

mixed model procedures can provide breeders with solutions for environmental effects that may be useful in management decisions.

The problems associated with breeding values estimated from contemporary group deviations were reemphasized by Willham and Leighton (1978). This method of breeding value estimation does not account for the genetic differences between groups and cannot eliminate the environmental differences between the groups. These problems led the researchers to propose a within herd mixed model evaluation that fit every animal in the herd. Willham and Leighton (1978) suggested the within herd animal model could be extended to tie all herds together and accomplish the goal of national sire evaluation to increase the number of sires that could be fairly compared from all sources of information.

Within herd BLUP procedures were further expanded by the development of a mixed model to obtain within herd predictors for additive direct and maternal weaning weight values (Slanger, 1977, 1979, and 1980). Slanger's model included all known genetic relationships among the animals, herd-year means, direct additive genetic values, maternal additive genetic values, and random environmental errors.

Willham (1979) pointed out some of the problems that were beginning to arise in beef sire evaluations. Superior sires were accumulating large numbers of progeny in many contemporary groups making it extremely difficult for young bulls with few progeny to come close to EPDs for the highly selected sires. This problem led to the inclusion of sire birth year groups into the mixed model. A sire relationship matrix was

later added to the mixed model equations to account for genetic trend (Berger, 1984).

Willham (1979) also pointed out the fear of some working with sire evaluations that breeders would begin to think that progeny tests were the only way to estimate sire breeding values. Use of an animal's own performance data and relative information has to be used also in sire evaluation to minimize generation interval. Experience gained in sire evaluation helped to identify evaluation weaknesses and stimulated the development of new ways to use all performance data, not just progeny data.

Quaas and Pollak (1980) demonstrated with BLUP procedures a methodology to use an animal's own performance data and to account for genetic correlation between traits being evaluated for sires. Their multiple trait model for weaning weight and yearling gain was called an animal model in contrast to a sire model because the equation for a record included a term for the breeding value of the animal making the record.

The expansion of mixed model methodology to include such things as multiple trait evaluations and an animal's own performance was complemented by other researchers working to integrate national sire evaluation and within herd evaluations. Middleton (1983) considered the integration of within herd BLUP and national sire evaluation as the best possible way to provide evaluation of within herd females and young animals. Middleton developed a within herd genetic merit adjustment factor to be applied to within herd predictors so that they

could be directly comparable to the base of the national sire evaluation. The adjustment factor was simply computed as the weighted average of sire EPDs used in the herd and was added to each within herd predictor. Simulation results verified that the adjustment factor was appropriate to account for different herd genetic merit levels.

Willham (1983) proposed a similar procedure for unifying within herd and national sire evaluations. Rather than doing independent within herd evaluations first and then applying the herd genetic merit adjustment factor, all within herd BLUP equations would be consolidated to form the national sire evaluation model. This would be done by absorbing within herd dam and contemporary group effects into herd sire equations that would be merged with other herd sire equations to form the national sire evaluation model. After solving for sire EPDs, back solutions for within herd predictors would be determined. There is no need to add the genetic merit adjustment factor with this method of determining within herd predictors. Within herd predictors would be automatically determined relative to the national sire evaluation base. It is this method of unifying national sire evaluation and within herd evaluations that is addressed in the following chapters.

DATA DESCRIPTION

Methodology and computer algorithms proposed for the unification of within herd and national sire evaluations are validated with actual weaning weight (WWT) field records. The records are provided by the courtesy of the American Angus Association, St. Joseph, Missouri.

A set of 20 herds is selected to form the validation data base. The herds generally represent long histories of performance testing and are tied by several sires through artificial insemination. A total of 30,413 individual WWT performance records is included in the data base spanning collection years from 1972 to 1983.

Table 1 presents a summary of the herd characteristics in terms of numbers of WWT records, contemporary groups, dams, and sires. Each record represents an individual animal that has recorded WWT performance data. Data fields included for each record are herd identification, sex, weaning date, weaning management code (creep or noncreep), designed test or field data, 205-day and age-of-dam adjusted weaning weight, sire registration, dam registration, and dam date of birth. A contemporary group for WWT is defined as the concatenation of data contained in the following fields: herd identification, date weaned (year and day), sex, weaning management code, and source of data (designed test or field).

The American Angus Association provided pedigree information for every sire with a progeny performance record as well as the performance field records. The pedigree data are used to construct the inverse of Wright's (1922) numerator relationship matrix (A) between sires.

Table 1. Summary of herd characteristics for weaning weight

Herd code	Within herd numbers			
	WWT records	Contemporary groups	Dams	Sires
1	1,274	26	363	39
2	2,275	58	760	66
3	3,154	153	1,195	90
4	898	16	378	21
5	467	28	158	42
6	3,035	141	980	129
7	1,513	45	529	61
8	1,336	74	540	26
9	897	25	343	62
10	2,333	30	780	80
11	991	55	364	73
12	1,095	115	394	68
13	953	54	301	86
14	139	6	71	14
15	301	49	110	21
16	4,296	21	1,459	60
17	2,842	56	822	73
18	340	37	166	20
19	2,201	40	782	98
20	71	20	35	17
Total	30,413	1,049	10,530	(850)[a]

[a]This is the total number of sires used across all herds with progeny records in the data, not the column sum.

The relationships used are a bull's sire and his maternal grandsire (MGS). Tables 2 and 3 summarize the structure of many of the sire relationships.

A total of 132 sires is added to the evaluation by virtue of having at least two male descendants (sons or grandsons, or any combination) with progeny records in the evaluation. The total number of sires in the evaluation is 982.

Table 2. Distribution of bulls with no progeny records of their own but with sons and grandsons in the sire evaluation[a]

Number		Number	
Bulls	Sons	Bulls	Grandsons
37	1	29	1
24	2	48	2
10	3	18	3
5	4	6	4
1	6	4	5
1	10	1	8
		1	13
		1	>20

[a]The number of bulls in each column are not mutually exclusive.

Table 3. Distribution of bulls with progeny records and sons and grandsons in the sire evaluation

Number		Number	
Bulls	Sons	Bulls	Grandsons
73	1	59	1
44	2	18	2
20	3	15	3
9	4	9	4
11	5	2	5
8	6	1	6
2	7	1	7
1	8	1	10
1	10	1	12
1	12	1	13
2	13	1	14
1	15	1	15
2	17	1	17
1	>20		

Relationships are extensive between the sires in the evaluation, adding strength to ties that already exist due to direct or indirect contemporary group ties. The increased number of sire ties resulting from including the relationships is evidenced by the fact that 499 bulls in the evaluation also have both a sire and maternal grandsire also in the evaluation with progeny records. Five hundred forty-six bulls in the evaluation receive added ties from being sons of bulls also having progeny records in the data. Another 280 bulls have added ties from being grandsons of bulls with progeny in the data. Many other ties are created also by the 132 sires that do not have progeny records in the evaluation.

Birth years for the sires are from 1964 to 1980. Eleven bulls with no birth year are automatically included in the 1964 year group for the analysis. Table 4 displays the distribution of the other 850 sires, excluding additions due to relationships.

A program to absorb dam effects into within herd contemporary group effects and sire effects is developed in support of this thesis. This program assumes that a dam can have only one progeny per contemporary group. However, several dams in the data do have twins in the same contemporary group. One of the twins is deleted from consideration in the validation effort. The total number of deleted twin records is 67. An adjustment to the WWT record of the remaining twin is warranted but is not done for this thesis. Future research needs to address how to handle the occurrence of twins in the same contemporary group.

Table 4. Distribution of bulls with progeny in the data by birth year of the bull

Birth year	Bulls	Cumulative percent (%)	Birth year	Bulls	Cumulative percent (%)
64	17	2.0	73	65	41.9
65	8	2.9	74	70	50.1
66	14	4.6	75	71	58.5
67	16	6.5	76	69	66.6
68	30	10.0	77	73	75.2
69	35	14.1	78	84	85.1
70	50	20.0	79	67	92.9
71	57	26.7	80	49	100.0
72	64	34.2			

The size of the validation data set dictates that the evaluation computational and storage tasks need to be comparable to those that would be required for an operational unified national beef sire evaluation and within herd evaluations. The mixed model evaluation equations and computer algorithms developed for this thesis can be used with minor modifications to handle a major breed's complete performance record data base. One of the modifications would be to allow for storage of various data sets on magnetic tape rather than on disk as is done in the validation process.

Within herd evaluations and sire evaluations are currently run independently of each other. The next chapter develops the mixed model evaluation methodology that is proposed for unifying beef sire and within herd evaluations.

METHODS AND PROCEDURES

This chapter begins by describing an operational evaluation implementation and timing concept for unifying national beef evaluation (NSE) and within herd evaluations. The first step requires the collection and forwarding of field records to a central processing center. The next step in implementation uses best linear unbiased prediction (BLUP) evaluation procedures and the resulting expected progeny differences (EPDs) of sires to determine within herd dam and young animal EPDs for various traits in each herd. The cycle is completed when breeders have access to new EPDs for use in selection decisions.

Definition of evaluation methodology follows the description of the evaluation implementation and timing concept. The evaluation methodology section develops BLUP equations to predict sire, dam, and young animal EPDs for direct and maternal weaning weight (WWT). The following section discusses how to compute and interpret within herd phenotypic, genetic, and environmental trends. The discussion explains how genetic trend lines can be computed from weighted EPDs of sires and dams that are used in the herd of a breeder. In addition, the section describes use of actual performance data and estimates of within herd contemporary group fixed effects to construct phenotypic and environmental trend lines, respectively.

Next, a computational procedures section outlines four major computer algorithms used in setting up sire evaluation equations and back solving for within herd environmental effects and dam and young

animal EPDs. Appropriate variance component ratios are key elements of BLUP procedures used in this thesis. The last section of the chapter, variance component estimation, discusses methods to estimate the variance components required for different evaluation models.

Evaluation Implementation and Timing Concept

Old performance data form the major foundation of seedstock evaluation when using BLUP procedures. However, the most recent performance data are essential for the evaluation of young breeding stock. This requires timely reporting of all new data to insure that breeding selection decisions are not made on old performance evaluation data or no data at all.

Effective and timely evaluation of seedstock within the purebred beef breeding industry requires the coordinated efforts of both breeders and evaluators. Performance measurements must be made and reported to the evaluation center as soon as possible after collection. Similarly, the evaluation center must analyze the records and return the appropriate evaluation data to the breeder within a matter of days.

The evaluation implementation and timing concept presented in this section is designed with the two requirements in mind. First, the implementation of sire and within herd evaluation procedures must match beef cattle production management schedules. Secondly, the procedures must be timed so that results are available before parent seedstock selection decisions are made.

The collection and evaluation processes seem to be formidable tasks even with modern large capacity, high speed computers.

Evaluation of an entire breed requires analyzing hundreds of thousands of records collected from across the country from two to three thousand different herds. Diversity of management schedules followed by breeders compounds the task of obtaining a quick turnaround. In addition, several economically important traits requiring analysis are not measured at the same time. For example, the first measurement on an animal occurs at birth, and the last measurement normally occurs at the end of a yearling gain period. Therefore, the time required for obtaining one generation's birth and growth performance records spans more than a year.

An ideal evaluation scheme could be done on a "real-time" basis. That is, immediately upon measurement of a trait, the identification of the animal and the measurement could be fed into a breeder's microcomputer storage buffer. After the buffer is full, or at a point when no more measurements are to be made, the buffer would transfer the data via an electronic communications network to a central processing computer. The data would be incorporated into a continuously running animal evaluation computer model after data verification, editing, and appropriate adjustments. Differences in breeding and management schedules would not affect this method of evaluation. Updated animal EPDs for various traits would become available as soon as new data from across the country were fed into the system. "Real-time" read outs of animal EPDs could be immediately available to breeders using personal remote microcomputer terminals.

In the ideal evaluation described, the genetic evaluation information obtained by a breeder would be more perishable than the current situation. New information would be constantly fed into the evaluation model rather than entered on a yearly basis as is now the case. Published reports would no longer be required. If a breeder wanted a customized listing for review, the breeder would merely ask for a listing of animals meeting some index criteria to be printed at a terminal.

The ideal evaluation described is possible with current technology. However, software development time and funding for computer hardware limit the industry from moving the current evaluation procedure to the ideal evaluation.

The proposed method of unifying within herd evaluations and NSE is a modest step in the direction of the ideal evaluation. The process does not approach "real-time," but does include two annual NSEs. The intent is to have the evaluations completed and available to breeders at key selection decision times. Various within herd evaluations would be done for dams and young animals following each NSE. The following paragraphs describe an operational timeline for a unified NSE and within herd evaluations.

Spring evaluations

Figure 1 shows the integration of performance evaluations with the different management aspects of a typical beef cattle breeding scenario. The schedule does not mean to imply that all calves are born in the months of February through April. However, a majority of the

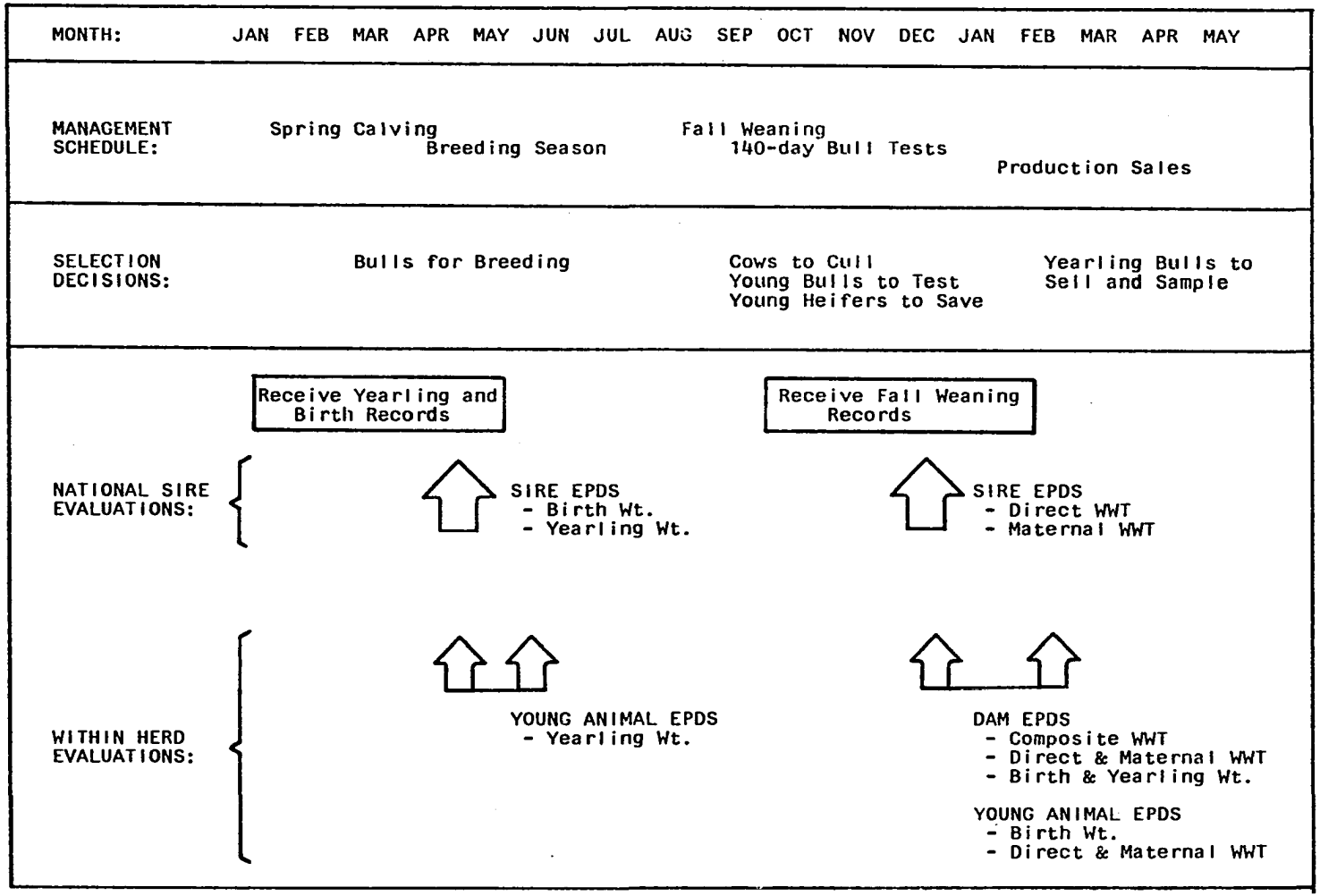

Figure 1. Schedule of unified national sire evaluation and within herd evaluations

calves are born during this time period. Until the ideal evaluation model is developed and implemented, evaluation procedures must be structured to aid a majority of the breeders.

The latest birth records from spring calving along with yearling weight records for calves born the prior year would be included in a spring NSE. The evaluation needs to be completed by the first of May. This insures that results are available before the major breeding season starts in late May and early June. Computing within herd young animal EPDs for yearling weight would follow the spring NSE. Herds would be evaluated with results sent back to breeders on a priority schedule tied to the date that their yearling and birth records were received.

Winter evaluations

A winter NSE would be accomplished in the middle of December, incorporating a majority of the current fall WWT records. Sire EPDs for direct and maternal WWT would be evaluated. A major within herd evaluation would follow the NSE during the months of January and February. The within herd evaluation would determine the following evaluation information:

1) Within herd phenotypic, genetic, and environmental trend charts,
2) dam EPDs for a composite direct and maternal WWT effect,
3) dam EPDs for direct and maternal WWT,
4) dam EPDs for birth and yearling weight, and
5) young animal EPDs for birth weight and direct and maternal WWT.

The effectiveness and success of a unified NSE and within herd evaluations concept will depend upon breeders and evaluators equally. Breeders must take and report animal trait measurements in a timely fashion. Similarly, evaluators must accomplish the evaluation immediately following receipt of birth and yearling performance records in the spring and after WWT performance records are received in the fall.

The method of disseminating evaluation results will play a key role in the success of the concept described. Breeders must have access to the evaluation results via a telephone modem and microcomputer network linked to a central processing center. Two benefits of such a network would be the following:

1) The breeder could have the information in a matter of minutes after the evaluation is completed, and

2) The breeder would not have to sort through thousands of sires to find the ones meeting his selection criteria.

Printed sire summaries for the large major breeds contain only sires evaluated with a moderately high accuracy. One reason for doing this is to reduce the volume of sires to be reported. However, if genetic progress is to be made by selection of parent seedstock, then listing only high accuracy sires is counter productive. The best genetic material for continued progress lies within the young sire gene pool having lower accuracy due to fewer progeny.

Some breed associations do prepare supplemental listings for a limited number of young sires. These lists are available to breeders upon request. However, a computer network would allow breeders to obtain

individualized sire, dam, and/or young animal listings for any young or old animals. Only animals meeting certain criteria of the breeder would be included in the listings. In addition, the animals could be ranked according to a breeder specified index.

The implementation of a unified sire and within herd evaluation requires the development of new evaluation methodology. The next section presents one approach to this unification.

Evaluation Methodology

The current method of constructing mixed model sire equations to determine sire expected progeny differences (EPDs) for direct weaning weight (WWT) is to absorb contemporary group fixed effects from all herds into the sire equations (Berger, 1984). A variance-covariance structure that includes sire relationships and the ratio of error variance to sire variance is added to the absorbed set of sire equations. The sire equations are then augmented with a set of group equations defined by birth year of sire. The equations are further augmented with a restriction in the form of a Lagrangian multiplier to correct for the linear dependency between contemporary group and sire birth year equations. Birth year groups are formed directly from the sire equations. However, no dependency exists between the group and sire equations, because the sire equations have a covariance matrix added after forming the birth year equations.

The Lagrangian multiplier chosen for the restriction does not need to be unique. A multiplier commonly chosen forces the summation of all birth year group solutions to zero. This restriction results in what

is termed a "floating base" sire evaluation. More records and sires are added to the evaluation from one year to the next. In addition, another sire birth year group will be added with the same restriction that all birth year group solutions sum to zero. New performance data and another birth year group cause the relative base of sire comparisons to change, hence the term "floating base." Relative differences between birth year group solutions are minor. However, these differences plus new sire estimates are enough to yield EPDs that are not comparable from one evaluation to the next. Each subsequent yearly evaluation determines new EPDs that contain the average increase in the mean of the breed. Only these values need to be used by the breeder when making selection decisions. The old values become irrelevant.

An alternative to a "floating base" is a "fixed base" evaluation. The Lagrangian multiplier is chosen to force a particular sire birth year group to the same value year after year. This birth year group then becomes the base of reference. The only change in a sire's EPD from one evaluation to the next is primarily due to the incorporation of additional progeny information on himself or on his relatives. As the number of progeny records for a sire becomes large, the EPD will stabilize and remain unchanged from then on.

One problem associated with the "fixed base" evaluation is that in the presence of positive genetic progress, the mean of all sire EPDs gets larger. With this comes a steady progression of all active sire EPDs into the positive range. There is a connotation associated with being positive versus being negative: "Even though a bull may be

ranked low, if his EPD is positive then he must not be too bad." This results in continued use of a bull beyond his genetic usefulness.

The proposed unification of NSE and within herd evaluations assumes a "floating base" evaluation. The ramifications of this assumption are not known at this time. Actual experience may dictate switching to a fixed base.

One method proposed for integrating within herd evaluations of a breed with the NSE involves modifying the basic mathematical model to include dam genetic effects in addition to sire genetic effects (Willham, 1983). It is this method that forms the basis for the evaluation methodology. Using this method, the sire equations used in the NSE would have dam effects as well as contemporary group effects included in the absorption process. The absorption would be on a herd by herd basis with dam and contemporary group equations stored for within herd back solution estimations. The final sire equations would be of the same form used in current sire evaluation procedures.

Back solutions for contemporary group fixed effects and dam effects would be made for each herd after solving for sire direct effects. The best linear unbiased estimates (BLUEs) for the fixed effects and the best linear unbiased predictors (BLUPs) for the sire and dam effects form the basis for determining other predictors such as: sire EPD for maternal WWT, dam EPDs for direct and maternal WWT, and young animal EPDs for direct and maternal WWT. The estimators and predictors would also be used to compute within herd genetic and environmental trends.

The following sections outline a proposed methodology for unifying within herd and national sire evaluations. Four basic evaluation models are developed for the proposed method of unification. Each of the four is derived from one general mathematical model for WWT.

General mathematical model definition

The WWT model shown in Figure 2 is patterned after the model of Willham (1963) in his study of the covariance between relatives when a trait is composed of two or more compound parts. The model forms the basis for all the evaluation models described in this thesis.

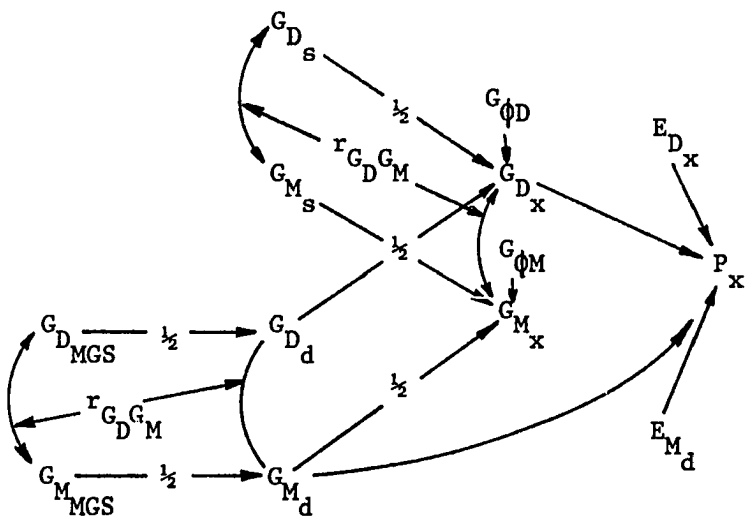

Figure 2. A path diagram describing a phenotypic value for weaning weight

The mathematical statement for the model in Figure 2 is given as follows:

$$P_x = G_{D_x} + E_x$$

where

G_{D_x} = the direct effect of x's genotype,

E_x = the total environmental influence on animal x,

and where E_x can be progressively partitioned as:

$$P_x = G_{D_x} + (G_{M_d} + E_{M_d} + E_{D_x})$$

where

G_{M_d} = the maternal effect on x caused by the genotype of the animal's dam (d),

E_{M_d} = a temporary environmental influence on the maternal value of the dam, and

E_{D_x} = an environmental influence on the direct effect of x.

Similarly, G_{D_x} can be partitioned as follows:

$$P_x = (1/2\ G_{D_s} + 1/2\ G_{D_d} + G_{\phi D}) + (G_{M_d} + E_{M_d} + E_{D_x})$$

where

G_{D_s} = the direct effect of the animal's sire (s),

G_{D_d} = the direct effect of the animal's dam (d), and

$G_{\phi D}$ = a genetic effect due to Mendelian sampling.

Genetic and environment effects making up P_x are rearranged to give an expression as follows:

$$P_x = \text{Fixed effect} + (1/2\ G_{D_s}) + (1/2\ G_{D_d} + G_{M_d})$$

$$+ (G_{\phi D} + E_{M_d} + E_{D_x} - \text{fixed effect})$$

$$= \text{Fixed effect} + \text{sire effect} + \text{dam effect} + \text{error} \qquad (1)$$

where the fixed effect includes the population mean, a sex effect, management effects, and herd-year-season effects. Some researchers have partitioned the dam's genetic maternal ability (G_{M_d}) into the components $G^*_{M_d}$ and E_{P_d}, where E_{P_d} represents a permanent environmental influence of the maternal effect of the dam (Quaas and Pollak, 1980).

Sire EPDs for direct WWT

Mathematical model The mathematical model given by Equation 1 can be rewritten for individual progeny in different herds to form the basic sire evaluation model as follows:

$$y_{ijklm} = u + c_{ij} + g_n + s_{nk} + d_{il} + e_{ijklm} \qquad (2)$$

where

y_{ijklm} = a 205-day adjusted WWT record for the m-th progeny of the l-th dam in the i-th herd and the k-th sire and reared in the j-th contemporary group of the i-th herd,

u = the population mean,

c_{ij} = a fixed effect common to each WWT record in the j-th contemporary group of the i-th herd,

g_n = a fixed effect common to each sire born in the n-th year,

s_{nk} = a random genetic effect for direct WWT associated with the k-th sire born in the n-th year, $\sim(0,\sigma^2_s A)$,

d_{il} = a random composite effect common to the l-th dam of the i-th herd, NID $(0,\sigma^2_d)$, and

e_{ijklm} = a residual random effect of the WWT record of m-th progeny born to the l-th dam and k-th sire, NID $(0,\sigma^2_e)$.

The random composite dam effect includes one-half her genetic effect for direct WWT ($1/2\ G_{D_d}$), her maternal effect for WWT (G_{M_d}), and a permanent environmental effect associated with her maternal effect (E_{P_d}) if using the partitioning of Quaas and Pollak (1980). A dam's permanent environmental effect is more difficult to estimate with a high degree of accuracy than it is to define and diagram mathematically (Figure 2). A dam's permanent environmental effect is the difference in actual capacity to produce milk and mother a calf and her genetic potential to do these things. Mastitis or frozen teats are examples of cases in which a cow's milking capacity might be impaired in the current and all future lactations.

For the purposes of this thesis, the permanent environmental effect is not differentiated from the genetic maternal effect in the prediction equations for a dam's EPD for maternal WWT. No attempt is made to define the methodology for estimating dam permanent environmental effects. If a permanent environmental effect exists for a cow, it is confounded with her genetic maternal effect (G_{M_d}).

The residual random effect (e_{ijklm}) includes a temporary environmental effect on the dam's maternal effect (E_{M_d}), an environmental influence on the m-th progeny's genetic direct effect for WWT, a Mendelian sampling effect, and all other errors associated with the

measurement of the WWT record for the progeny. The fixed effect (c_{ij}) includes sex, herd-year-season, and management effects.

Mixed model Ignoring sire birth year effects for the present, the expression given by Equation 2 can be represented in matrix notation to include all records as follows:

$$\underline{y} = X\underline{c} + Z_1 \underline{u}_1 + Z_2 \underline{u}_2 + \underline{e} \qquad (3)$$

where

$\underline{y}$ = a known vector of 205-day adjusted WWT records,

X = a known incidence matrix associated with contemporary groups,

$\underline{c}$ = a vector of unknown contemporary group fixed effects,

Z_1 = a known incidence matrix defining dams of the progeny expressing the WWT record,

$\underline{u}_1$ = a vector of unknown dam random effects $\sim(0,\sigma_d^2 \cdot I)$,

Z_2 = a known incidence matrix defining sires of the progeny expressing the WWT record,

$\underline{u}_2$ = a vector of unknown sire random effects for direct WWT $\sim(0,\sigma_s^2 \cdot A)$, and

$\underline{e}$ = a vector or uncorrelated residual errors, NID $(0,\sigma_e^2)$.

The mixed model equations for the model given by Equation 3 are:

$$\begin{bmatrix} X'X & X'Z_1 & X'Z_2 \\ Z_1'X & \begin{pmatrix} Z_1'Z_1 & Z_1'Z_2 \\ Z_2'Z_2 & Z_2'Z_2 \end{pmatrix} & + \; G^{-1} \end{bmatrix} \begin{bmatrix} \tilde{\underline{c}} \\ \tilde{\underline{u}}_1 \\ \tilde{\underline{u}}_2 \end{bmatrix} = \begin{bmatrix} X'\underline{y} \\ Z_1'\underline{y} \\ Z_2'\underline{y} \end{bmatrix} \qquad (4)$$

where

$$G = 1/\sigma_e^2 \; \text{Var}\,(\underline{u})$$

$$= 1/\sigma_e^2 \left[\begin{array}{c|c} \sigma_d^2 \cdot I & 0 \\ \hline 0 & \sigma_s^2 \cdot A \end{array}\right] .$$

The matrix A is Wright's (1922) numerator relationship matrix defining the covariances among the sires represented in the model.

Currently, computational feasibility of the mixed model equations given by Equation 4 necessitates the assumption that dams are unrelated. Absorbing two effects (contemporary groups and dam effects) to form the national sire evaluation equations requires that two inverses be computed. If both matrices (X'X and $Z_1'Z_1$) are diagonal, it is numerically preferred to invert and absorb the largest matrix first. The inverse of a diagonal matrix can be computed directly by replacing each diagonal element with one divided by the element. In most herds, $Z_1'Z_1$ will be of a much larger order than X'X and will only be diagonal if relationships among dams are ignored. Considering the fact that as many as 2500 herds may be evaluated together and that many herds will have in excess of 500 dams (active and inactive), a monumental computational inversion task is involved if other than a diagonal matrix for dams is used.

Evaluation process The process of absorbing dams and contemporary groups to form national sire evaluation equations can be best explained by the following series of steps:

Step 1: Sort all performance data by herd and dams within herd.

Step 2: Absorb dams directly into contemporary group and sire equations (within the i-th herd) to form the following matrices:

$$\begin{bmatrix} X'SX & X'SZ_2 \\ Z_2'SX & Z_2'SZ_2 \end{bmatrix}_{\text{i-th herd}}, \qquad \begin{bmatrix} X'S\underline{y} \\ Z_2'S\underline{y} \end{bmatrix}_{\text{i-th herd}}$$

where

$$S = I - Z_1(Z_1'Z_1 + \sigma_e^2/\sigma_d^2 \cdot I)Z_1'.$$

Step 3: Compute $(X'SX)^{-1}$ and absorb contemporary groups into the sire equations to form:

$$(Z_2'S^*Z_2)_{\text{i-th herd}} \text{ , } (Z_2'S^*\underline{y})_{\text{i-th herd}}$$

where

$$S^* = S - SX(X'SX)^{-1} X'S.$$

Step 4: Concurrently with Steps 2 and 3, elements of the matrices $(X'SX)^{-1}$, $X'S\underline{y}$, and $X'SZ$ must be stored for back solution of within herd fixed effects ($\hat{\underline{c}}$) and dam effects ($\hat{\underline{u}}_1$).

Step 5: Repeat Steps 2-4 for each herd.

Step 6: Sort and sum the coefficient elements of $(Z_2'S^*Z_2)$ and $(Z_2'S^*\underline{y})$ by sire ID to form the equations for the national sire evaluation.

Step 7: Augment the sire equations with G^{-1}, the variance-covariance matrix, and the sire birth year groups to form the following mixed model equations:

$$\begin{bmatrix} T'S^*T & T'S^*Z_2 & \underline{1} \\ Z_2'S^*T & Z_2'S^*Z_2 + A^{-1}\cdot\sigma_e^2/\sigma_s^2 & 0 \\ \underline{1} & 0 & 0 \end{bmatrix} \begin{bmatrix} \tilde{\underline{g}} \\ \tilde{\underline{u}}_2 \\ LM \end{bmatrix} = \begin{bmatrix} T'S^*\underline{y} \\ Z_2'S^*\underline{y} \\ 0 \end{bmatrix} \qquad (5)$$

where

T = a known incidence matrix defining the birth year group to which a sire belongs,

$\tilde{\underline{g}}$ = an unknown vector of birth year fixed effects, and

LM = Lagrangian multiplier.

Step 8: Iterate the equations of Equation 5 to determine sire direct effect predictors ($\hat{\underline{u}}_2$) and birth year group estimates ($\tilde{\underline{g}}$). The best linear unbiased predictor of EPD for direct WWT for the k-th sire born in the j-th year is given by:

$$\hat{u}^*_{nk} = \hat{g}_n + \hat{u}_{nk}. \quad (6)$$

Back solutions The BLUEs for contemporary group fixed effects for the i-th herd can be determined with the matrices found in Step 2. The solutions are given by the following equation:

$$\underline{\hat{c}}_i = (X'SX)^{-1}_i((X'S\underline{y})_i - (X'SZ_2)_i \underline{\hat{u}}^*_2) \quad (7)$$

where

$\underline{\hat{u}}^*_2$ = a known vector or sire EPDs for direct WWT.

The composite direct and maternal effect for each dam within the i-th herd can be found by back solution of Equation 4. The first step in the back solution process is to substitute known contemporary group fixed effects ($\underline{\hat{c}}$) and known sire EPDs ($\underline{\hat{u}}^*_2$) for direct WWT into Equation 4. The solutions are given by the following equation:

$$\underline{\hat{u}}_{1i} = ((Z_1'Z_1 + \sigma^2_e/\sigma^2_d \cdot I)^{-1} ((Z_1'\underline{y}) - (Z_1'X)\underline{\hat{c}}_i - (Z_1'Z_1) \underline{\hat{u}}^*_2) \quad (8)$$

where

$\underline{\hat{u}}_{1i}$ = an unknown vector of within herd dam effects for WWT,

$\underline{\hat{c}}_i$ = a known vector of within herd contemporary group fixed effects, and

$\underline{\hat{u}}^*_2$ = a known vector of sire EPDs for direct WWT.

Variance ratios Current estimates for the variance components that are required for beef sire evaluation when including dam effects do not exist in the literature. One approach for estimating the variance components is discussed in the section entitled "Variance Component Estimation."

Accuracy The goal of national beef sire evaluations is to increase the number of sires that can be fairly compared on breeding

value differences obtained from all sources of information (Willham, 1974). Procedures for fair comparisons include: regressing predicted breeding values back to a breed (or genetic group) average depending upon the amount of information available for evaluation; regressing the predicted breeding value for incomplete heritability; accounting for the distribution of sire progeny across contemporary groups; and accounting for average genetic merit of the sires of the progeny in a contemporary group. Accounting for these factors allows all sire breeding values to be directly and fairly compared.

One measure associated and reported with a sire's predicted breeding value or expected progeny difference is accuracy. Accuracy can be considered as a reflection of the amount that a sire's EPD has been regressed back to the breed average. A low accuracy value indicates that the regression is considerable. Low accuracy values indicate few progeny records from which to evaluate the sire. High accuracy values result from large amounts of information available from which to evaluate the sire. The heritability (h^2) of the trait and the number of progeny by a sire influence accuracy. As h^2 decreases, more progeny are necessary to achieve a fixed level of accuracy. Accuracy can be defined statistically as a function of the squared correlation between a sire's predicted EPD and his true EPD and can be derived from the known properties of the mixed model (Berger, 1984).

Accuracies of EPDs have been represented in several ways: effective progeny number, possible change, percent of perfect accuracy, or just accuracy ($R^2_{\hat{u}u}$). Regardless of the representation, accuracy is directly

related to prediction error variance (PEV), or the variance of the difference between a sire's predicted EPD and his true EPD. PEV can be obtained from the generalized inverse of Henderson's (1973) mixed model coefficient matrix. The PEV for any sire's EPD is the inverse diagonal element associated with the sire in question times the residual error variance of the evaluation model.

The coefficient matrix in many sire evaluations is of such a large order that computing its generalized inverse is cost prohibitive. Solutions for sire EPDs in these cases are obtained by an iterative procedure for solving simultaneous equations, and PEVs to be used in accuracy calculations must be approximated. Most approximations are a function of the inverse of the diagonal element of the coefficient matrix for the sire in question.

Ufford et al. (1979) developed an approximation procedure for determining prediction error variances in dairy sire evaluations using all lactation records in a BLUP procedure. Simple regression was used to relate prediction error variance obtained from the actual coefficient matrix inverse to the diagonal element of the coefficient matrix after absorption of cow, sire-by-year, natural service sire, and herd-year-season equations. The researchers cautioned against using their regressions with evaluations done under different models. Ufford et al. (1979) further suggested that when genetic groups are included in the evaluation model and selection is among sires in different groups, an appropriate measure of accuracy would be:

$$R^2_{\hat{s}s} = 1 - \frac{\text{Var } (\hat{g} + \hat{s} - s)}{\text{Var } (s)} \qquad (9)$$

where Var $(\hat{g} + \hat{s} - s)$ is the error variance of the evaluation and Var (s) is sire variance for the trait of the evaluation.

The 1983 beef sire evaluations conducted at Iowa State University define prediction accuracy as follows:

$$R^2_{\hat{s}s} = 1 - \frac{\text{Var } (\hat{s} - s)}{\text{Var } (s)}$$

$$\cong 1 - (\sigma^2_e/\sigma^2_s)/D \qquad (10)$$

where D is the diagonal element of $(Z_2'S^*Z_2 + G^{-1})$. This diagonal element contains a contribution from relationships that will increase its magnitude. However, the contribution from relatives with few progeny records will create an upward bias in the magnitude of D. Therefore, the relationship contribution is adjusted by an amount proportionate to the number of progeny records of each relative in the evaluation. The adjustment is used to create a more conservative approximation for PEV.

Prediction error variance is approximated since the true inverse elements are impossible to obtain from large sets of equations. The approximation assumes all other sires compared with the sire in question are estimated with perfect accuracy. Since many other sires are predicted with less than perfect accuracy, the approximation over-estimates the true accuracy. An adjustment which considers the number of progeny along direct lines of descent is given as follows:

$$R^2_{\hat{s}s} \cong 1 - (\alpha/(EPN_i + K_1\alpha - K_2))$$

where

$$\alpha = \sigma^2_e/\sigma^2_s,$$

K_1 = the diagonal element of the relationship inverse, and

K_2 = a relationship adjustment factor accounting for the number progeny records of close relatives.

This computable form relies on a new approximation for prediction error variance given by:

$$\text{Var}\ (\hat{s}_i - s_i) \cong \left[\frac{1}{EPN_i + a_{ii}\alpha - k_i}\right]\sigma_e^2$$

where

EPN_i = the effective number of progeny for the i-th sire,

a_{ii} = the diagonal inverse element of relationships for the i-th sire,

$\alpha = \sigma_e^2/\sigma_s^2$, and

k_i = an adjustment factor depending on the relationship between the i-th bull and his own sire and/or sires with progeny records in the evaluation.

The adjustment factor k_i is given by the following expression:

$$k_i = (.06667\alpha)^2\left[\sum_{j=1}^{n_s} \frac{1}{EPN_j + a_{jj}\alpha - k_j} + \frac{1}{EPN_k + a_{kk}\alpha - k_k}\right]$$

where

n_s = the number of sires of the i-th bull with progeny records in the evaluation,

EPN_j = the effective number of progeny for the j-th sire of the i-th bull, and

EPN_k = the effective number of progeny for the sire of the i-th bull.

A subset of the validation herds is used to form a smaller set of sire mixed model evaluation equations than when using the complete set of herds. Twenty herds are used and the evaluation equations contain 982 sires and 18 birth year groups. The coefficient matrix for these sires and groups is small enough to invert directly to obtain PEV for

a sire's predictor. The actual PEV is compared to the approximation given by Equation 10. A summary of the findings is presented in the "Results and Discussion" chapter of this thesis.

A sire EPD for direct WWT constitutes one part of the sire evaluation for the weaning weight trait. The other part is the sire's EPD for maternal WWT. The next part of this section discusses a proposed method of determining sire EPDs for maternal WWT.

Sire EPDs for maternal WWT

Mathematical model A sire's expected progeny difference (EPD) for maternal WWT can only be obtained by analyzing WWT records of progeny from female relatives. The method developed in this paper considers only the progeny records of daughters of a sire in determining his maternal EPD.

Consider the mathematical model statement for a calf's WWT record given by the following expression:

$$y_{ijklm} = u + c_{ij} + s_k + d_{il} + e_{ijklm} \qquad (11)$$

where

y_{ijklm} = a 205-day adjusted WWT record of the m-th progeny of the k-th sire and the l-th dam in the i-th herd and reared in the j-th contemporary group of the i-th herd,

u = the population mean,

c_{ij} = a fixed effect common to each WWT records made in the i-th contemporary group of the j-th herd,

s_k = a random genetic effect for direct WWT associated with the k-th sire $\sim(0,\sigma_s^2 \cdot A)$,

d_{il} = a random genetic composite effect for direct and maternal WWT associated with the l-th dam of the i-th herd $\sim(0, \sigma_d^2 \cdot I)$, and

e_{ijklm} = a residual random effect in the record of the m-th progeny of the k-th sire mated to the l-th dam of the i-th herd and in the j-th contemporary group of the i-th herd. This error term includes a temporary environmental effect on the dam's maternal effect, NID(0,σ^2_e).

It is assumed that contemporary group estimates ($\hat{c}_{ij}$) and sire predictors ($\hat{s}^*_k$) for direct WWT are available from the within herd/ national sire evaluation. The model is identical to Equation 2 with the birth group effect considered as part of the sire effect. An adjusted WWT record is given by:

$$(y_{ijklm})_{adjusted} = y_{ijklm} - \hat{c}_{ij} - \hat{s}^*_k \qquad (12)$$

where

$\hat{c}_{ij}$ = the estimated environmental fixed effect for the j-th contemporary group in the i-th herd,

$\hat{s}^*_k$ = the EPD for the k-th sire.

The adjusted WWT record is a function of the progeny's dam effect (d_{il}) and the random error component (e_{ijklm}). The dam effect includes both a genctic direct effect and a maternal effect for WWT and can be written as follows:

$$d_{il} = 1/2\ G_{D_d} + G_{M_d} + E_{P_d} \qquad (13)$$

where

G_{D_d} = the breeding value for direct WWT for the l-th dam in the i-th herd,

G_{M_d} = the breeding value for maternal WWT for the l-th dam in the i-th herd,

E_{P_d} = a permanent environmental effect associated with the maternal effect of the l-th dam in the i-th herd.

The dam's breeding values (G_{D_d}, G_{M_d}) can be written either as a function of both her sire and dam breeding values or as a function only of her sire with her dam's breeding value expressed as a random error effect. This sire is the maternal grandsire (MGS) of the progeny expressing the adjusted record. The substitution of MGS breeding value for dam breeding value is given by the following expression:

$$(y_{ijklm})_{adjusted} = 1/2\ (1/2\ G_{D_{MGS}}) + 1/2\ G_{M_{MGS}} + E_{P_d} + e'_{ijklm} \qquad (14)$$

where

e'_{ijklm} = error term which includes the maternal granddam's contribution not accounted for and an error term for the calf, e_{ijklm}.

Substitution of transmitting abilities for breeding values in Equation 14 results in:

$$(y_{ijklm})_{adjusted} = 1/2\ S_{D_{MGS}} + S_{M_{MGS}} + e_{ijklm} \qquad (15)$$

where

$S_{D_{MGS}}$ = one-half of the MGS's breeding value for direct WWT,

$S_{M_{MGS}}$ = one-half of the MGS's breeding value for maternal WWT,

e_{ijklm} = a random error effect that includes both E_{P_d} and e' of Equation 14.

Mixed model equations The expression given by Equation 15 can be written in matrix notation to accommodate all progeny records as follows:

$$\underline{y} = Z_1\underline{u}_1 + Z_2\underline{u}_2 + \underline{e} \qquad (16)$$

where

$\underline{y}$ = a known vector of 205-day WWT records adjusted for the environmental fixed effect ($\hat{c}_{ij}$) and the sire's direct effect predictor ($\hat{s}^{*}_{k}$),

Z_1, Z_2 = known incidence matrices relating MGSs to records of daughters with Z_1 associated with the direct effect for WWT and containing 0's and .5's, Z_2 is associated with the maternal effect for WWT and contains 0's and 1's,

$\underline{u}_1$, $\underline{u}_2$ = unknown vectors of random effects for direct and maternal WWT, respectively. They are bivariate normal with $E(\underline{u}_i)=0$, and

$\underline{e}$ = a vector of residual random errors, potentially correlated due to record adjustment, but assumed NID $(0, \sigma^2_e)$.

The random model equations used to solve for sire EPDs and direct and maternal WWT are given by:

$$\begin{bmatrix} \begin{pmatrix} Z_1'Z_1 & Z_1'Z_2 \\ Z_2'Z_1 & Z_2'Z_2 \end{pmatrix} + G^{-1} \end{bmatrix} \begin{bmatrix} \tilde{\underline{u}}_1 \\ \tilde{\underline{u}}_2 \end{bmatrix} = \begin{bmatrix} Z_1'\underline{y} \\ Z_2'\underline{y} \end{bmatrix} \tag{17}$$

where

$$G^{-1} = \text{Var}\,(\underline{u})^{-1} \cdot \sigma^2_e$$

$$= \left[\begin{array}{c|c} \sigma^2_{s_D} \cdot I & \sigma_{s_D s_M} \cdot I \\ \hline \sigma_{s_D s_M} \cdot I & \sigma^2_{s_M} \cdot I \end{array}\right]^{-1} \sigma^2_e \tag{18}$$

and where

σ^2_e = residual random error variance,

$\sigma^2_{s_D}$ = sire variance for direct WWT,

$\sigma^2_{s_M}$ = sire variance for maternal WWT, and

$\sigma_{s_D s_M}$ = sire covariance for direct and maternal WWT.

It is assumed that dams are unrelated in the proposed within herd/ national sire evaluation model for computational purposes only. No relationships are accounted for among the MGSs due to this assumption. Therefore, the identity matrix (I) in Equation 18 is used rather than the inverse of Wright's (1922) numerator relationship matrix (A^{-1}). With the four identity matrices, G^{-1} is a tridiagonal matrix. The tridiagonal structure and a possible rearrangement into a block diagonal matrix (G^{*-1}) are illustrated for three sires as follows:

$$G^{-1} = \begin{array}{c} \\ (1) \\ (2) \\ (3) \\ (1) \\ (2) \\ (3) \end{array} \begin{array}{c} \begin{array}{cccccc} (1) & (2) & (3) & (1) & (2) & (3) \end{array} \\ \begin{bmatrix} \sigma^2_{s_D} & 0 & 0 & \sigma_{s_D s_M} & 0 & 0 \\ 0 & \sigma^2_{s_D} & 0 & 0 & \sigma_{s_D s_M} & 0 \\ 0 & 0 & \sigma^2_{s_D} & 0 & 0 & \sigma_{s_D s_M} \\ \sigma_{s_D s_M} & 0 & 0 & \sigma^2_{s_M} & 0 & 0 \\ 0 & \sigma_{s_D s_M} & 0 & 0 & \sigma^2_{s_M} & 0 \\ 0 & 0 & \sigma_{s_D s_M} & 0 & 0 & \sigma^2_{s_M} \end{bmatrix}^{-1} \sigma^2_e \end{array}$$

$$G^{*-1} = \begin{array}{c} \begin{matrix} (1) & & (2) & & (3) & \end{matrix} \\ \left[\begin{array}{cc|cc|cc} \sigma^2_{s_D} & \sigma_{s_D s_M} & 0 & 0 & 0 & 0 \\ \sigma_{s_D s_M} & \sigma^2_{s_M} & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & \sigma^2_{s_D} & \sigma_{s_D s_M} & 0 & 0 \\ 0 & 0 & \sigma_{s_D s_M} & \sigma^2_{s_M} & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & \sigma^2_{s_D} & \sigma_{s_D s_M} \\ 0 & 0 & 0 & 0 & \sigma_{s_D s_M} & \sigma^2_{s_M} \end{array}\right]^{-1} \end{array} \sigma^2_e .$$

The coefficient matrix of Equation 17 can be rearranged into a block diagonal matrix taking advantage of the tridiagonal structure. The rearranged matrix is given by the following expression:

$$\begin{bmatrix} P_1'P_1 + \sigma^2_e \text{ Var } (\tilde{\underline{b}}_1)^{-1} & & & \\ & P_2'P_2 + \sigma^2_e \text{ Var } (\tilde{\underline{b}}_2)^{-1} & & \\ & & \ddots & \\ & & & P_n'P_n + \sigma^2_e \text{ Var } (\tilde{\underline{b}}_n)^{-1} \end{bmatrix} \begin{bmatrix} \tilde{\underline{b}}_1 \\ \tilde{\underline{b}}_2 \\ \vdots \\ \tilde{\underline{b}}_n \end{bmatrix} = \begin{bmatrix} P_1'\underline{y} \\ P_2'\underline{y} \\ \vdots \\ P_n'\underline{y} \end{bmatrix} \quad (19)$$

where

P_i = an incidence matrix which defines all of the progeny records belonging to daughters of the i-th MGS,

$\tilde{\underline{b}}_i$ = a 2x1 vector of $\tilde{\underline{u}}_{1_i}$ and $\tilde{\underline{u}}_{2_i}$ for the i-th MGS, and

$$\text{Var } (\tilde{\underline{b}}_i)^{-1} = \begin{bmatrix} \sigma^2_{s_D} & \sigma_{s_D s_M} \\ \sigma_{s_D s_M} & \sigma^2_{s_M} \end{bmatrix}^{-1}.$$

Birth year of MGS group equations are added to Equation 19 to account for genetic trend in maternal WWT effect. The grouping is consistent with the grouping criteria used in the sire evaluation for direct WWT effect. After adding groups, Equation 19 becomes:

$$\begin{bmatrix} T'T & T'P \\ P'T & P'P \oplus \sigma^2_e \text{ Var } (b)^{-1} \end{bmatrix} \begin{bmatrix} \tilde{\underline{g}} \\ \tilde{\underline{b}} \end{bmatrix} = \begin{bmatrix} T'\underline{y} \\ P'\underline{y} \end{bmatrix} \qquad (20)$$

where

T = a known incidence matrix that assigns MGSs to their respective birth year groups, and

$\tilde{\underline{g}}$ = an unknown vector of birth year group fixed effects.

The equations in Equation 20 are of full rank and require no restriction prior to inverting or iterating in order to find the solutions for $\tilde{\underline{g}}$ and $\tilde{\underline{b}}$. Prior estimated predictors from the sire evaluation for direct WWT should be used in the equations in Equation 20. The prior estimates are determined with a smaller prediction error variance than those that would be found by solving Equation 20.

Variance components Appropriate estimates of sire genetic maternal effect for WWT and its covariance with sire genetic direct effect for WWT do not exist in the literature. See pages 91 to 99

for a discussion of methods to determine variance and covariance associated with direct and maternal effects for WWT.

Prediction accuracy Prediction accuracy for a sire's EPD for maternal WWT is calculated in the same manner as the prediction accuracy for direct WWT. The formulation of maternal EPD accuracies is derived from the mixed model equations of Equation 20. Following the procedures of Henderson (1973), the variance of the model effects is given by:

$$\text{Var}\begin{bmatrix}\tilde{g}\\ \tilde{b}\end{bmatrix} = \begin{bmatrix}T'T & T'P\\ P'T & P'P \oplus \sigma_e^2 \cdot \text{Var}\ (b)^{-1}\end{bmatrix}^{-1} \cdot \sigma_e^2$$

$$= \begin{bmatrix}C^{11} & C^{12}\\ C^{21} & C^{22}\end{bmatrix} \cdot \sigma_e^2 \quad . \tag{21}$$

The prediction error variance associated with the i-th sire is given by:

$$\text{Var}\ (\hat{b}_i - b) = C_{ij}^{22} \cdot \sigma_e^2$$

$$\cong (P_i'P_i + \sigma_e^2\ \text{Var}\ (b)^{-1})^{-1} \cdot \sigma_e^2. \tag{22}$$

The variance of the vector of direct and maternal effects is given by:

$$\text{Var}\ (b) = \begin{bmatrix}\sigma_{s_D}^2 & \sigma_{s_D s_M}\\ \sigma_{s_D s_M} & \sigma_{s_M}^2\end{bmatrix} \tag{23}$$

so that

$$\frac{\text{Var } (\hat{b}_i - b)}{\text{Var } (b)} = (P_i'P_i + \sigma_e^2 \cdot \text{Var } (b)^{-1})^{-1} \begin{bmatrix} \sigma^2_{s_D} & \sigma_{s_D s_M} \\ \sigma_{s_D s_M} & \sigma^2_{s_M} \end{bmatrix}^{-1} \sigma_e^2$$

$$= \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} \sigma_e^2. \qquad (24)$$

Using the diagonal element of Equation 24, the prediction accuracy for a sire's EPD for maternal WWT is given by:

$$R^2 \cong 1 - D_{22} \cdot \sigma_e^2. \qquad (25)$$

Sire EPDs for maternal WWT would be presented along with sire EPDs for direct WWT in the proposed national sire summary. Both direct and maternal EPDs would be reported in the same format as + or - XX lbs. Accuracies would be given on a scale from 0 to 1. The next part addresses methodology to determine within herd dam EPDs for individual herd owners or for listing top dams of the breed.

Dam EPDs for direct and maternal WWT

Knowledge of EPDs for individual dams, under most circumstances, is not as important to a breeder as knowing the EPDs for potential service sires. Each dam will rarely have more than one progeny in a given calf crop. In contrast, a sire used in natural service may be used with 20 or more females. Sires being used artificially may be used on hundreds of cows. Therefore, the overall genetic merit of a calf crop is heavily influenced by a few males and each cow in the herd.

Little selection pressure can be applied on the female side of a beef breeding herd. Culling old cows and cows with bad udders requires

the breeder to keep back a large percentage of heifers. Culling for genetic merit of females with progeny requires saving an even larger percentage of heifers. The mean breeding value becomes less for every heifer kept for breeding and approaches the mean breeding value of the parent generation assuming that replacement heifers are not chosen at random. Therefore, selection pressure must come from the male side of the breeding population for genetic improvement to occur.

Breeders have two situations in which an EPD for direct WWT and an EPD for maternal WWT on dams are desirable. The first case is a breeder that considers buying females from another herd, a common practice for a young breeder who is building a foundation cow herd. Secondly, an EPD for direct WWT and an EPD for maternal WWT are valuable information to a breeder who wants to superovulate a cow for embryo transfers.

A dam's composite effect for direct and maternal WWT is easily obtainable from back solution of within herd equations with knowledge of contemporary group fixed effects and sire EPDs for direct WWT. Determining a separate dam EPD for direct WWT and one for maternal WWT is more involved and is developed in the following paragraphs.

Mathematical model Referring back to the original evaluation model given by Equation 2, a 205-day adjusted WWT record is composed of a contemporary group fixed effect (c_{ij}), a sire direct effect for WWT (s_{nk}), a dam composite effect for direct and maternal WWT (d_{il}), and a residual error. After solving for sire EPDs for direct WWT and back solving for within herd contemporary group fixed effects ($\hat{c}_{ij}$),

the WWT record can be adjusted for these estimated effects leaving a random dam effect (d_{il}) and an error term. The adjustment is written as:

$$y_{ijklm} - \hat{c}_{ij} - \hat{s}^*_{nk} = d_{il} + e_{ijklm} \tag{26}$$

where

$$d_{il} = 1/2\ G_{D_d} + G^*_{M_d} + E_{P_d} \tag{27}$$

where

G_{D_d} = breeding value for direct WWT for the l-th dam of the i-th herd,

$G^*_{M_d}$ = breeding value for maternal WWT for the l-th dam of the i-th herd,

E_{P_d} = permanent environmental influence on the maternal effect of the l-th dam in the i-th herd.

The permanent environmental influence is confounded with the dam's genetic maternal effect. It would be difficult to determine best linear unbiased predictors for a dam's permanent environmental effect without good parameter estimates for its variance and covariance with the other effects. In addition, not all dams will have such an effect. To partition a part of the dam's WWT effect into permanent environment does not appear to offer the breeder any useful genetic evaluation information. The partitioning requires use of variance components to allocate a portion to $G^*_{M_d}$ and a portion to E_{P_d} (Quaas and Pollak, 1980).

It can be argued that if a permanent environmental effect exists, then it should be a part of the dam's genetic maternal effect without distinction. Such an argument is particularly true if the susceptibility to mastitis or other udder and teat anomalies is passed on to her

offspring (Willham, Department of Animal Science, Iowa State University, 1984). Following this argument, the permanent environmental effect is combined with the dam's genetic maternal effect for the remainder of the discussion with no attempt at estimating it. Substituting the dam composite effect given by Equation 27 into Equation 26 gives:

$$(y_{ijklm})_{adjusted} = 1/2\ G_{D_d} + G_{M_d} + e_{ijklm}. \tag{28}$$

All progeny adjusted records can be represented in matrix notation as follows:

$$\underline{y} = Z_1\ \underline{u}_1 + Z_2\ \underline{u}_2 + \underline{e} \tag{29}$$

where

Z_1 = a known incidence matrix of 0's and .5's,

Z_2 = a known incidence matrix of 0's and 1's,

$\underline{u}_1$ = an unknown vector of dam breeding values for direct WWT,

$\underline{u}_2$ = an unknown vector of dam breeding values for maternal WWT, and

$\underline{e}$ = vector of residual random effects, NID$(0,\sigma^2_e)$.

The random model equations for Equation 29 are given by:

$$\begin{bmatrix} \begin{pmatrix} Z_1'Z_1 & Z_1'Z_2 \\ Z_2'Z_1 & Z_2'Z_2 \end{pmatrix} + G_1^{-1} \end{bmatrix} \begin{bmatrix} \tilde{\underline{u}}_1 \\ \tilde{\underline{u}}_2 \end{bmatrix} = \begin{bmatrix} Z_1'\underline{y} \\ Z_2'\underline{y} \end{bmatrix} \tag{30}$$

where

$$G_1^{-1} = \left[\begin{array}{c|c} \sigma^2_{d_D} \cdot I & \sigma_{d_D d_M} \cdot I \\ \hline \sigma_{d_D d_M} \cdot I & \sigma^2_{d_M} \cdot I \end{array}\right]. \tag{31}$$

The variances and covariances of Equation 31 are defined as follows:

$\sigma^2_{d_D}$ = dam variance for direct WWT, this value is assumed equal to sire variance for direct WWT,

$\sigma^2_{d_M}$ = dam variance for maternal WWT, this value is assumed equal to sire variance for maternal WWT, and

$\sigma_{d_D d_M}$ = covariance between dam direct and maternal WWT, also assumed to be equal to sire covariance for direct and maternal.

The equations of Equation 30 can be rearranged into a set of independent dam equations because effects for each dam are assumed to be uncorrelated with those of any other dam. Each dam will have two equations and two unknowns, one for her direct WWT EPD and one for her maternal WWT EPD. An example set of equations for the 1-th dam of the i-th herd is given as follows:

$$\left[\begin{pmatrix} n_{il} & n_{il} \\ n_{il} & n_{il} \end{pmatrix} + \sigma^2_e \begin{pmatrix} \sigma^2_{d_D} & \sigma_{d_D d_M} \\ \sigma_{d_D d_M} & \sigma^2_{d_M} \end{pmatrix}^{-1}\right] \begin{bmatrix} \tilde{u}_{1_{il}} \\ \tilde{u}_{2_{il}} \end{bmatrix} = \begin{bmatrix} \Sigma y_{ADJ_{il}} \\ \Sigma y_{ADJ_{il}} \end{bmatrix} \quad (32)$$

where

n_{il} = the number of progeny records for the 1-th dam of the i-th herd,

$\tilde{u}_{1_{il}}$ = EPD for direct WWT for the 1-th dam in the i-th herd,

$\tilde{u}_{2_{il}}$ = EPD for maternal WWT for the 1-th dam in the i-th herd, and

$\Sigma y_{ADJ_{il}}$ = the summation of all adjusted WWT records as given by Equation 12 for the 1-th dam in the i-th herd.

Variance components The variance components required for the dam direct and maternal WWT model are the same as those required for the sire EPD for maternal WWT model. See pages 90 to 91 for a discussion of variance components.

Prediction accuracy The accuracy of prediction for the dam direct and maternal EPDs will be low due to the low number of progeny records for each dam. The prediction error variance (PEV) will be the diagonal element of the inverse of the coefficient matrix given in Equation 32. Therefore, prediction accuracy is defined as:

$$R^2_{\hat{u}_{1_{il}} \, u_{1_{il}}} = 1 - \frac{\text{Var}\,(\hat{u}_{1_{il}} - u_{1_{il}})}{\text{Var}\,(u_{1_{il}})}$$

$$= 1 - \frac{PEV}{\sigma^2_{d_D}} . \qquad (33)$$

Similarly,

$$R^2_{\hat{u}_{2_{il}} \, u_{2_{il}}} = 1 - \frac{PEV}{\sigma^2_{d_M}} . \qquad (34)$$

One way to improve prediction accuracy is to include information from a dam's sire. The dam's sire will have EPDs for direct and maternal WWT that are BLUPs obtained from progeny and grand-progeny records across herds, respectively. Incorporation of a dam's sire information can be similar to a procedure given by Henderson (1975a) to include animals in mixed model equations even though they have no progeny records. Henderson's purpose was to increase accuracy of sire evaluations by increasing the number of covariance ties between sires.

The situation with the dam direct and maternal WWT model differs from Henderson's model in some respects. First, prior estimates of predictors for the animals without progeny records will be known. Secondly, each animal added is assumed to be related to only one dam in the model. These differences allow for the consideration of a set of independent pairs of equations, one for each dam in the herd. The dam's sire equations are added directly to each set as defined in Equation 32. For the l-th dam of the i-th herd, the augmentation can be expressed as:

$$\left[\begin{array}{cc|c} n_{il} & n_{il} & \\ & & \phi \\ n_{il} & n_{il} & \\ \hline & \phi & \phi \end{array} + G_{2_{il}}^{-1}\right]\left[\begin{array}{c} \tilde{u}_{1_{il}} \\ \tilde{u}_{2_{il}} \\ \hat{u}^{*}_{1_k} \\ \hat{u}^{*}_{2_k} \end{array}\right] = \left[\begin{array}{c} Z'_{il}\underline{y} \\ Z'_{2l}\underline{y} \\ 0 \\ 0 \end{array}\right] \qquad (35)$$

where

$\hat{u}^{*}_{1_k}$ = sire of dam EPD for direct WWT,

$\hat{u}^{*}_{2_k}$ = sire of dam EPD for maternal WWT,

and where

$$G_{2_{il}}^{-1} = \sigma_e^2 \text{ Var} \left[\begin{array}{c} u_{1_{il}} \\ u_{2_{il}} \\ \hat{u}^{*}_{1_k} \\ \hat{u}^{*}_{2_k} \end{array}\right]^{-1}$$

$$= \sigma_e^2 \left[\begin{array}{cc|cc} 1\cdot\sigma^2_{s_D} & 1\cdot\sigma_{s_D s_M} & .5\cdot\sigma^2_{s_D} & .5\cdot\sigma_{s_D s_M} \\ 1\cdot\sigma_{s_D s_M} & 1\cdot\sigma^2_{s_M} & .5\cdot\sigma_{s_D s_M} & .5\cdot\sigma^2_{s_M} \\ \hline .5\cdot\sigma^2_{s_D} & .5\cdot\sigma_{s_D s_M} & 1\cdot\sigma^2_{s_D} & 1\cdot\sigma_{s_D s_M} \\ .5\cdot\sigma_{s_D s_M} & .5\cdot\sigma^2_{s_D} & 1\cdot\sigma_{s_D s_M} & 1\cdot\sigma^2_{s_M} \end{array}\right]^{-1}$$

$$= \sigma_e^2 \left[\begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix} \otimes \begin{pmatrix} \sigma^2_{s_D} & \sigma_{s_D s_M} \\ \sigma_{s_D s_M} & \sigma^2_{s_M} \end{pmatrix}\right]^{-1}$$

$$= \sigma_e^2 \begin{bmatrix} 1 & .5 \\ .5 & 1 \end{bmatrix}^{-1} \otimes \begin{bmatrix} \sigma^2_{s_D} & \sigma_{s_D s_M} \\ \sigma_{s_D s_M} & \sigma^2_{s_M} \end{bmatrix}^{-1}. \qquad (36)$$

EPDs for young animals may be more important to breeders than dam EPDs for direct and maternal WWT. Young animals collectively carry the genetic potential for the greatest breed improvement. The young animals with the greatest breeding potential need to be identified as soon as possible so that their genes can be incorporated into the parent gene pool. The next part of this section discusses a proposed method to obtain EPDs for direct and maternal WWT for the young animal.

Young animal EPDs for direct WWT

Breeding value estimates for animals produced in the latest generation are important for breed improvement. A breeder must choose heifers that will replace cows being culled. Young bulls to be sampled or offered for sale must also be identified and ranked according to their genetic merit.

The breeding values for young animals need to be directly comparable to as many animals as possible to increase a breeder's intensity of selection. A best linear unbiased predictor (BLUP) for a young animal's expected progeny difference (EPD), or one-half its breeding value, is the type of predictor that allows direct EPD comparisons between animals. This assumes that animals to be compared are contained within the same BLUP evaluation model.

The breeding value of an individual animal is equal to the sum of the average effects of the genes it carries (Falconer, 1960). The expected breeding value is one-half the sum of the breeding values for the parents, because every animal carries a sample half of the genes from each parent. However, the breeding value is likely to be different from the expected value when considering an isolated animal.

Modern day animal geneticists accept the concept that transfer of genetic information from parent to offspring follows a set of laws set forth by Gregor Mendel in the 19th century. Pairs of genes in each parent segregate to form gametic cells that recombine to form a zygote with a new genotype. The genotype, while similar to the parents, will be different. It is the random nature of this segregation and

recombination that causes full sibs to be genetically dissimilar (Lush, 1948). Some researchers refer to this phenomena as the Mendelian sampling effect in the formation of new genotypes (Quaas and Pollak, 1980).

Mathematical model One linear predictor of a young animal's EPD for direct WWT would include the animal's expected value based upon the parent EPDs for direct WWT plus a best linear unbiased predictor for the animal's Mendelian sampling effect. The following is an expression for a young animal's EPD for direct WWT:

$$EPD_{animal} = 1/2\ (1/2\ G_{D_s} + 1/2\ G_{D_d} + G_{\phi D}) \tag{37}$$

where

G_D = an animal's breeding value for direct WWT,

$G_{\phi D}$ = genetic effect from Mendelian segregation and recombination.

Parent (sire and dam) EPDs for direct WWT ($1/2\ G_{D_s}$ and $1/2\ G_{D_d}$) are available from the national sire evaluation and within herd dam evaluation, respectively. Formulation of a method to determine $G_{\phi D}$ begins with the following expression:

$$P_x = G_{D_x} + E_x \tag{38}$$

where

P_x = a WWT performance record of the animal adjusted for the population mean and contemporary group effect,

G_{D_x} = breeding value for the animal, and

E_x = environmental influence on the animal's genotypic expression.

The adjusted performance record given by Equation 38 can be reexpressed as follows:

$$y_{ijklm} - (\widehat{u+c}_{ij}) = (1/2\ G_{D_s} + 1/2\ G_{D_d} + G_{\phi D}) + (G_{M_d} + E_{P_d} + e) \quad (39)$$

where

y_{ijklm} = a 205-day adjusted WWT record for the m-th progeny of the l-th dam and k-th sire and reared in the j-th contemporary group of the i-th herd,

$(\widehat{u+c}_{ij})$ = population mean plus a fixed effect common to all records in the j-th contemporary group of the i-th herd,

G_{D_s}, G_{D_d} = breeding values for direct WWT for the k-th sire and l-th dam in the i-th herd, respectively,

G_M = breeding value for maternal WWT for the l-th dam in the i-th herd,

$G_{\phi D}$ = Mendelian sampling effect for the m-th progeny,

E_{P_d} = permanent environmental influence on the maternal effect of the l-th dam in the i-th herd, and

e = a residual random effect.

Consideration of all deviated progeny records can be represented in matrix notation as follows:

$$\underline{y} = Z_1\underline{u}_1 + Z_2\underline{u}_2 + Z_3\underline{u}_3 + \underline{e} \quad (40)$$

where

$\underline{y}$ = a vector of deviated progeny WWT records,

Z_i = known incidence matrices relating progeny, dams, and sires to each record, respectively,

$\underline{u}_1$ = an unknown vector of Mendelian sampling effects for each young animal,

$\underline{u}_2$ = a vector of dam random composite effects for direct and maternal WWT,

$\underline{u}_3$ = a vector of sire random effects for direct WWT, and

e = a vector of the residual random error effects.

Predictors for $G_{\phi D}$ The random model equations for determining BLUPs for $G_{\phi D}$ are given by:

$$\begin{bmatrix} Z_1'Z_1 + \sigma_e^2/\sigma_\phi^2 \cdot I & Z_1'Z_2 & Z_1'Z_3 \\ Z_2'Z_1 & Z_2'Z_2 + \sigma_e^2/\sigma_d^2 \cdot I & Z_2'Z_3 \\ Z_3'Z_1 & Z_3'Z_2 & Z_3'Z_3 + \sigma_e^2/\sigma_s^2 \cdot A^{-1} \end{bmatrix} \begin{bmatrix} \tilde{\underline{u}}_1 \\ \tilde{\underline{u}}_2 \\ \tilde{\underline{u}}_3 \end{bmatrix} = \begin{bmatrix} Z_1'\underline{y} \\ Z_2'\underline{y} \\ Z_3'\underline{y} \end{bmatrix}. \quad (41)$$

If prior estimates are available for $\tilde{\underline{u}}_2$ and $\tilde{\underline{u}}_3$, then $\tilde{\underline{u}}_1$ can be solved directly as follows:

$$\hat{\underline{u}}_1 = (Z_1'Z_1 + \sigma_e^2/\sigma_d^2 \cdot I)^{-1}(Z_1'\underline{y} - Z_1'Z_2\hat{\underline{u}}_2 - Z_1'Z_3\hat{\underline{u}}_3) \quad (42)$$

where $G_{\phi D}$ for the m-th progeny in the j-th contemporary group of the i-th herd is given by:

$$\hat{G}_{\phi D} = \frac{1}{(1 + \sigma_e^2/\sigma_\phi^2)} (y_{ijklm} - \widehat{u+c}_{ij} - \hat{d}_{il} - \hat{u}^*_{jk}) \quad (43)$$

where

$\widehat{u+c}_{ij}$ = estimate of the j-th contemporary group effect in the i-th herd,

$\hat{d}_{il}$ = composite direct and maternal WWT effect for the l-th dam of the i-th herd, and

$\hat{u}^*_{nk}$ = EPD for direct WWT for the k-th sire born in the n-th year.

Variance components An approximation for σ_ϕ^2 can be found from Equation 37 as follows:

$$\text{Var}(G_{D_x}) = (1/2)^2 \text{Var}(G_{D_s}) + (1/2)^2 \text{Var}(G_{D_d}) + \text{Var}(G_{\phi D})$$

or

$$\sigma^2_{A_D} = 1/4\ \sigma^2_{A_D} + 1/4\ \sigma^2_{A_D} + \sigma^2_{\phi}$$

such that

$$\sigma^2_{\phi} = 1/2\ \sigma^2_{A_D}$$

$$= 1/2\ (4\ \sigma^2_s) \qquad (44)$$

where

$\sigma^2_{A_D}$ = additive genetic variance for direct WWT, and

σ^2_s = sire genetic variance for direct WWT (note: $\sigma^2_s = 1/4\ \sigma^2_{A_D}$).

The residual error variance (σ^2_e) in Equation 41 is not the same error variance of the sire and dam mixed model given by Equation 4 (defined here as $\sigma^2_{e'}$). The variance ratio to be used in Equation 41 is obtained from:

$$\sigma^2_{e'}/\sigma^2_s = \frac{\sigma^2_{\phi}+\sigma^2_e}{\sigma^2_s}$$

such that

$$\sigma^2_e/\sigma^2_{\phi} = 1/2\ \left[(\sigma^2_{e'}/\sigma^2_s)-2\right]. \qquad (45)$$

The resulting EPD for direct WWT for a young animal is given by:

$$EPD_{animal} = 1/2\ (EPD_{sire} + EPD_{dam}) + 1/2\ \hat{G}_{\phi D} \qquad (46)$$

where the EPD_{sire} and EPD_{dam} are for direct WWT also.

A study of Equation 46 shows that an animal's EPD is determined as a function of the performance of paternal and maternal half-sibs. That is, EPDs for the sire and dam are determined from their progeny performance records. The value of $G_{\phi D}$ is based upon an animal's own

performance. The value of $G_{\phi D}$ also determines the similarity or dissimilarity of an animal to the average of his half-sibs.

Determination of young animal EPDs for birth and yearling weight follows the same procedure given for direct WWT. First, a BLUP of the animal's $G_{\phi D}$ for the trait must be solved. This value is substituted into Equation 46 along with the animal's sire and dam EPDs for the trait.

A Mendelian sampling effect, $G_{\phi M}$, does not exist for maternal WWT for the young animal, because the performance record does not contain influence from the animal's own maternal effect (see Figure 2). Therefore, the EPD for maternal WWT for the young animal becomes one-half the sum of the parent EPDs for maternal WWT. An expression for this is given as follows:

$$EPD_{animal} = 1/2\ EPD_{sire} + 1/2\ EPD_{dam} \tag{47}$$

where

EPD_{animal} = a young animal's EPD for maternal WWT,

EPD_{sire} = the sire's EPD for maternal WWT, and

EPD_{dam} = the dam's EPD for maternal WWT.

Accuracy The PEV for the EPD of a young animal for direct WWT is a function of the PEVs for the three BLUPs used in Equation 41. Let the EPD for the young animal given in Equation 41 be expressed as $L'\hat{\underline{u}}$, where L is a linear vector and $\underline{u}$ is the vector of BLUPs. The prediction error variance of $L'\hat{\underline{u}}$ is given by:

$$\text{Var}\ (L'\hat{\underline{u}} - L'\underline{u}) = E\left[\left[(L'\hat{\underline{u}} - L'\underline{u}) - E(L'\hat{\underline{u}} - L'\underline{u})\right]\cdot \left[(L'\hat{\underline{u}} - L'\underline{u}) - E(L'\hat{\underline{u}} - L'\underline{u})\right]'\right]. \tag{48}$$

Using BLUP properties given by Henderson (1973) and the expectations that $\hat{\underline{u}}$ and $\underline{u}$ are zero, it can be shown that:

$$\text{Var}\ (L'\hat{\underline{u}} - L'\underline{u}) = L'\ (\text{Var}(\underline{u}) - \text{Var}\ (\hat{\underline{u}}))L$$

$$= L'(C_{22} \cdot \sigma_e^2)L \tag{49}$$

where

C_{22} = inverse of the coefficient matrix of Equation 41.

Using a definition of prediction accuracy, $R^2_{\hat{u}u}$, given by:

$$R^2_{\hat{u}u} = 1 - \frac{\text{Var}(L'\hat{\underline{u}} - L'\underline{u})}{\text{Var}(L'\underline{u})} \tag{50}$$

an approximation for $R^2_{\hat{u}u}$ becomes:

$$R^2_{\hat{u}u} \cong 1 - L'(\text{Diag}(Z'Z + G^{-1}))\ L\ G^{-1}\ \sigma_e^2. \tag{51}$$

The prediction accuracy for the young animal's EPD for direct WWT can be approximated by:

$$R^2_{\hat{u}u} \cong 1 - l_1^2(1 - \text{Acc.}_{sire}) - l_2^2(1 - \text{Acc.}_{dam})$$

$$- l_3^2(\alpha/(1+\alpha)) \tag{52}$$

where

l_i = 1/2, and

α = ratio of error variance to Mendelian sampling variance, $(\sigma_e^2/\sigma_\phi^2)$.

A series of phenotypic, genetic, and environmental trend lines can be constructed when EPDs are available for sire, dams, and young animals belonging to or having progeny in a given herd. The next section of this chapter addresses how to compute such trend lines.

Phenotypic, Genetic, and Environmental Trends

A graphic historical perspective using phenotypic, genetic, and environmental trend lines can convey a quick assessment of a breeder's success in previous generations. Trend lines may be used to compare alternative methods of selection or management for breeders involved with large herds. Trend lines may reinforce selection and management goals established by a breeder. However, such trend lines may also signal the need for changes in selection and/or management schemes. Trend lines tending in a direction and/or at a rate contrary to the anticipated result could mean that projected responses to selection were made with inappropriate assumptions or that expectations were too optimistic.

A trend plot could consist of plotting all calf weaning weights adjusted to 205 days for each year that such records have been collected. Alternatively, plotting arithmetic means for all the records for each year would consolidate the data for easier interpretation. Connecting the points on a plot of arithmetic means results in a phenotypic trend line. Least squares means of weaning weights for each year provide unbiased estimates for each year by adjusting for variability and unequal numbers from year to year. The linear regression line joining the least squares means could be used to provide an unbiased estimate of future phenotypic performance.

Results illustrated in a phenotypic trend are a combination of genetic and environmental influences. Therefore, two observations of a phenotypic trend are possible. First, the trend line given by the

least squares regression line is either positive, negative, or unchanging with time. Secondly, the amount of variability in the plot of yearly arithmetic means is either large, small, or somewhere in between. Knowledge of the data going into the trends may allow a breeder to interpret a trend line. However, it is likely that a breeder will not be able to distinguish between the contributions of genetic and environmental influences.

A breeder needs separate genetic and environmental trend lines in addition to the phenotypic trend line for adequate interpretation of the data. Genetic trend is defined as the change in production per unit of time due to change in mean breeding value. Environmental trend is defined as the change in production per unit of time due to change in mean environment (Harville and Henderson, 1967).

Much has been written about the estimation of genetic trends and environmental trends in breeding populations with emphasis placed on removing environmental differences in performance data to determine unbiased estimates of the genetic trend (selected references include: Van Vleck and Henderson, 1961; Smith, 1962; Harville and Henderson, 1967; Schaeffer et al., 1975; and Hintz et al., 1978). Environmental differences arise from sources such as age, sex, location, management, health, etc. However, a breeder's within herd perspective of environmental influences and genetic trend is somewhat different.

A successful breeder tries to provide his or her stock with an environment that allows the animals to fully express their genetic potential. Biological systems perform more efficiently under some

environmental conditions than under others. A breeder must decide what can be done to enhance the environment within the constraints of cost and expected genetic return.

Phenotypic trends

A plot of phenotypic trends for a herd requires taking, recording, and reporting measurements for all the animals. Records from only registered animals, for example, could bias arithmetic means and the least squares linear regression line.

Three different phenotypic trends on 205-day adjusted weaning weights of interest to breeders might include:

1) A phenotypic trend line determined from considering all weaning weight records (adjusted for age of dam and to 205 days) regardless of sex or rearing management differences. The trend line would consist of connecting arithmetic means calculated on a yearly basis as follows:

$$P_{ij} = (\sum_{k} y_{ijk})/n_{ij} \tag{53}$$

where

P_{ij} = the phenotypic average for WWT in the j-th year of the i-th herd,

y_{ijk} = the adjusted weaning weight record of the k-th calf in the j-th year of the i-th herd, and

n_{ij} = the total number of calves belonging to the j-th year of the i-th herd.

2) The same as (1), except that male calves are considered separately from female calves.

3) The same as (1) or (2), with fall born calves separated from spring born calves.

Genetic trends

The method of mixed model equations provides a powerful tool for estimating genetic and environmental trends (Henderson, 1973). The mixed model equations contain environmental fixed effects and random genetic effects that result in phenotypic expression. Appropriately weighted estimates for environmental fixed effects and sire predictors can be used to determine the within herd environmental trend and genetic trend, respectively.

The genetic merit of any calf group is a function of the genetic material supplied by the parents. Each progeny receives a sample half of the genes carried by the sire and a sample half carried by the dam. These genes recombine to form a genetically different, yet similar, animal. Considering all the progeny of a given calf crop, the expected genotypic average of the progeny (genetic merit) equals one-half the collective genotypic value (breeding value) of the sires and one-half the collective genotypic value of the dams.

The mixed model equations from the unified within herd and national sire evaluations provide estimates of one-half the breeding value for both sires and dams. The genetic merit of a calf crop can be estimated as a weighted average of the sires' expected progeny differences (EPDs) and the dams' EPDs having progeny in the calf crop. The EPDs are equal to one-half the sire's or dam's breeding value.

The genetic merit for direct WWT of a calf crop born in the j-th year of the i-th herd consisting of contributions from both sires and dams can be expressed as:

$$G_{ij} = (\sum_{k} n_{ijk} \hat{s}^{*}_{k})/n_{ij\cdot} + (\sum_{l} \hat{d}_{il})/n_{ij\cdot} \qquad (54)$$

where

G_{ij} = the genetic merit of the calf crop weaned in the j-th year of the i-th herd,

n_{ijk} = the number of progeny of the k-th sire in the j-th year of the i-th herd,

$\hat{s}^{*}_{k}$ = the EPD for direct WWT of the k-th sire,

$\hat{d}_{il}$ = the EPD for direct WWT of the l-th dam in the i-th herd with a progeny record in the j-th year, and

n_{ij} = the total number of progeny in the j-th year of the i-th herd.

A breeder may also be interested in the genetic merit of a cowherd expressed as a single function over time, similar to the way commercial breeders view their cow herds. A commercial breeder is likely to be interested in a composite genetic value for a cow that includes both direct and maternal effects with no distinction between the effects for two reasons. First, a commercial breeder is paid for pounds of calf weaned. The cow that weans the heaviest calf year after year is the best cow regardless of whether it was due to her direct genetic effect for weaning weight or due to her maternal effect. Secondly, selection emphasis by a breeder to improve the maternal genetic merit in replacement animals is accomplished best from proven sires with high maternal genetic merit.

A composite dam effect trend line can be developed in the following two ways:

1) Plot raw means of dam composite effect predictors for direct and maternal WWT ($\hat{d}_{il}$) estimated from the sire/dam within herd mixed model as a function of birth year of the dams. The means include both active and inactive dams. Superimpose on this plot least squares means to assess the overall trend. The regression line accounts for numbers of dams in each birth year and provides an unbiased estimate of future birth years assuming that no major shifts in selection goals have occurred.

2) Plot raw means of dam composite effect predictors as a function of birth year of dam for only the active females. That is, only consider those dams with progeny in the latest calf crop. This removes past historical information that may be of interest in assessing long-term trends. However, the plot gives a more appropriate reflection of the current parent females in the herd.

Environmental trends

Environmental influences upon traits of economic importance can arise from many sources. The milking ability of a dam is an environmental influence on the weaning weight of her calf. Weather is another environmental influence upon growth traits of growing animals over which the breeder has little control. Supplemental feeding is an example of an environmental influence that the breeder can control. Regardless of the origins, environmental influences are defined as the difference between the phenotypic expression and the genetic contribution.

Alternatively, environmental trend can be computed from the following expression:

$$E_{ij} = (\sum_{k} n_{ijk} \hat{c}_{jk})/n_{ij}. \quad (55)$$

where

E_{ij} = the environmental effect in the j-th year for the i-th herd,

n_{ijk} = the number of calves in the k-th contemporary group of the j-th year,

$\hat{c}_{jk}$ = the estimated effect for the k-th contemporary group of the j-th year, and

n_{ij} = the total number of progeny records in the j-th year for the i-th herd.

A successful breeder attempts to provide an environment that allows for the fullest expression of genetic potential. If genetic contribution from the parent stock is consistently increasing and performance stays the same, the environment needed for the fullest expression of genetic merit has not been provided. However, improving environment can cost money. The cost must be weighed against the added genetic expression that it allows.

Promotional herd trend chart

Merchandising seedstock is one of the most important tasks a breeder has. The genetic potential possessed in his/her herd has little value unless it can be profitably marketed. Herd genetic trend lines have potential for use by breeders in selling their seedstock based on merit.

A breed association conducting unified sire and within herd evaluations should establish one acceptable herd trend chart for use

by all breeders advertising in the breed's major journal. The herd trend chart would be prepared by an association during the annual within herd evaluation. The herd trend chart would include information subject to change since the chart is directly linked to a sire evaluation that has a floating base. The changing nature of herd trend charts would require each chart to carry a date of publication and an association seal that could not be altered.

A sample "Herd Trend Chart" for marketing purposes is shown in Figure 3. The chart has the breeder's name, the date of publication, and the association's seal of endorsement. The performance (P) and environmental (E) trend lines are given in actual weaning weight pounds. Contribution of sire direct effects (G_s) for weaning weight to each year's respective calf crop is given as a deviation from the breed's sire genetic trend for weaning weight. The breed's genetic trend is the solid linear regression line. The total genetic contribution from both sire direct effects and dam composite effects (G_T) to each year's calf crop is also shown as a deviation from the linear regression line.

The previous sections in the chapter address the proposed evaluation methodology in compact mixed model matrix notation. The actual computational processes generally do not involve direct matrix operations. Instead, a series of numerical processes are developed that provide the same end result in a different manner than direct matrix operations. The next section outlines four computational algorithms that were developed in support of this thesis.

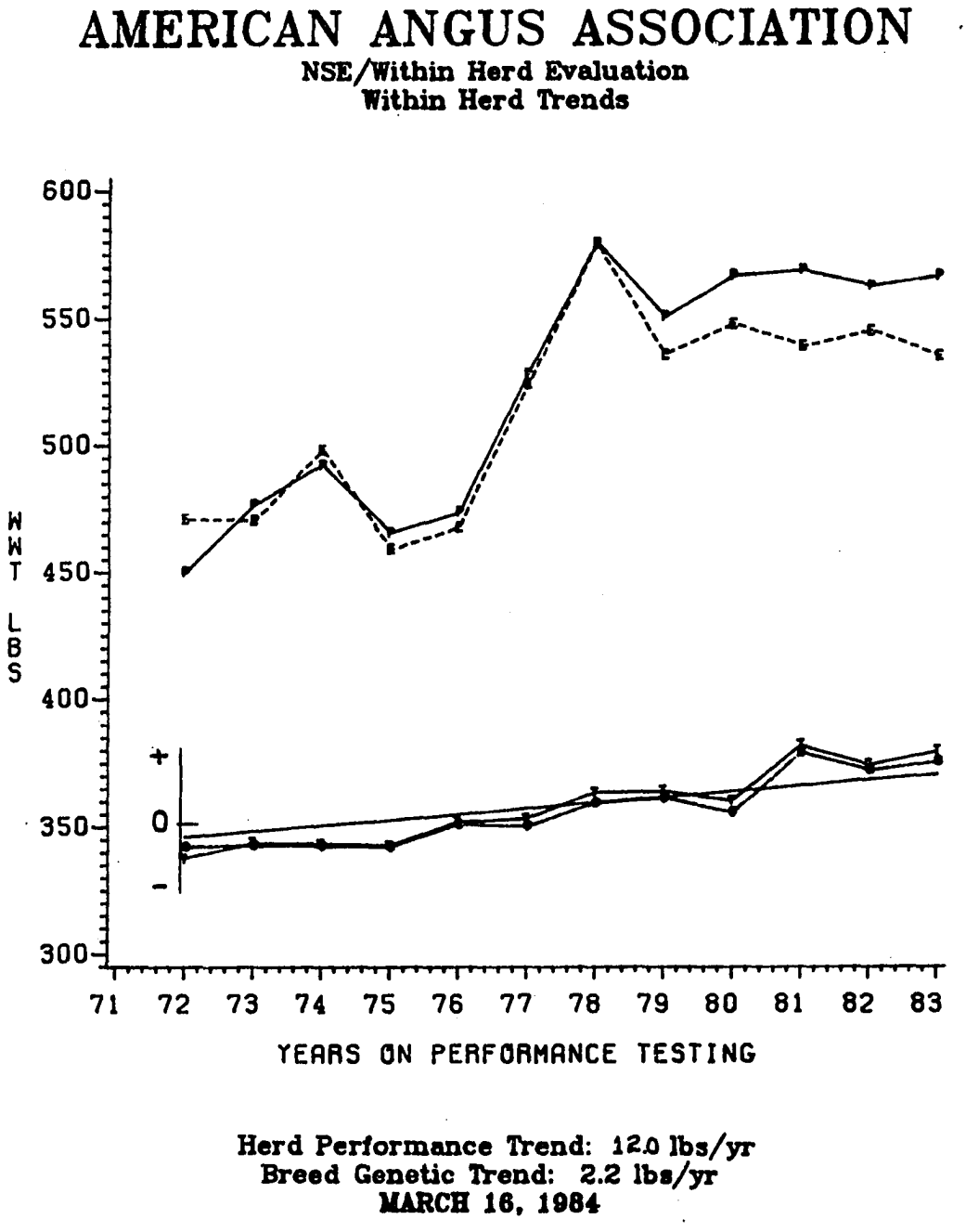

Figure 3. Sample herd trend chart

Computational Procedures

Computer algorithms developed to support the unification of within herd evaluations and national sire evaluations are written in PL/I (Programming Language 1). PL/I is a structured computer language that includes the file handling and formula solving capabilities necessary to efficiently perform animal breeding mixed model evaluations. The programs are compiled prior to execution using the PL/I Optimizing Compiler.

Basic descriptions of four algorithms presented in this thesis are as follows:

1) Absorption algorithm -- within herd dams,
2) Absorption algorithm -- within herd contemporary groups,
3) Back solutions -- contemporary group effects, and
4) Back solutions -- dam composite effects.

Absorption algorithm -- within herd dams

The purpose of the within herd dam absorption algorithm is to build directly within herd contemporary group and sire least squares equations with the dam effects absorbed as performance records are read into the computer. The PL/I program listing is contained in Appendix A. The absorption algorithm requires the dam matrix $(Z_1'Z_1 + G_{11}^{-1})$ to be diagonal. This matrix is diagonal when relationships between dams do not exist or are ignored. The dam by dam matrix is given in dot matrix notation as follows:

$$\begin{bmatrix} N_{i\cdot\cdot 1} + \alpha & & & \\ & N_{i\cdot\cdot 2} + \alpha & \phi & \\ & & \ddots & \\ \phi & & N_{i\cdot\cdot 1} + \alpha & \\ & & & \ddots \\ & & & N_{i\cdot\cdot n_d} + \alpha \end{bmatrix} \qquad (56)$$

where

$N_{i\cdot\cdot 1}$ = the total number of progeny records for the 1-th dam in the i-th herd,

α = the error variance to dam variance ratio, and

n_d = the total number of dams in the i-th herd.

Dam absorption is the first computational procedure to be done after the performance records have been edited and sorted by herds, by dams within herds, and by contemporary groups within dams. That is, all performance records of a given herd are sequenced together. In addition, all progeny records for a dam within a herd are sequenced together with the oldest record appearing first, followed by the next oldest, etcetera.

The coefficients for five different matrices must be determined during the dam absorption process. This includes coefficients for: $X'SX$, $X'SZ_2$, $X'Sy$, $Z_2'SZ_2$, and $Z_2'Sy$. As each progeny record is read for a given dam, a partial contribution to each coefficient is determined, labeled with a row and column index, coded to indicate the specific matrix with which it is associated, and written to a disk file. For example, row and columns indices for the $X'SX$ matrix are

contemporary group identifications. Row and column indices for $X'SZ_2$ are contemporary identification and sire registration number, respectively. After all partial contributions to each coefficient have been determined and written to the disk file, the file can be sorted by row and column and summed within each identical row and column combination to build the final coefficients for each of the five matrices. The matrices are used in the within herd contemporary group absorption process that is the next computational step.

Procedures to build directly the dam absorbed matrices can be best understood from a study of each matrix expressed in dot matrix notation. The matrices are presented following a general definition of selected variables and the dummy subscripts:

N_{ijkl} = 1 or 0. If the l-th dam and the k-th sire have a progeny in the j-th contemporary group and the i-th herd, then the value is 1, otherwise the value is 0.

$N_{i\cdot\cdot l}$ = the total number of progeny for the l-th dam in the i-th herd.

$N_{ij\cdot\cdot}$ = the total number of records in the j-th contemporary group of the i-th herd.

$N_{i\cdot kl}$ = the total number of records the l-th dam and the k-th sire have in common.

n_d = the total number of dams within the i-th herd.

n_c = the total number of contemporary groups in the i-th herd.

n_s = the total number of sires used within the i-th herd.

Matrix X'SX, contemporary group by contemporary group, is given in Equation 57.

$$X'SX = X'X - X'Z_1 \left(Z_1'Z_1 + G_{11}^{-1}\right) Z'X$$

$$= \begin{bmatrix}
N_{i1\cdot\cdot} - \sum_{l=1}^{n_d} N_{i1\cdot l}^2 \cdot \left(\frac{1}{N_{i\cdot\cdot l}+\alpha_l}\right) & -\sum_{l=1}^{n_d} N_{i1\cdot l}\cdot N_{i2\cdot l}\cdot\left(\frac{1}{N_{i\cdot\cdot l}+\alpha_l}\right) & \cdots\cdots & -\sum_{l=1}^{n_d} N_{i1\cdot l}\cdot N_{in_c\cdot l}\cdot\left(\frac{1}{N_{i\cdot\cdot l}+\alpha_l}\right) \\
-\sum_{l=1}^{n_d} N_{i2\cdot l}\cdot N_{i1\cdot l}\cdot\left(\frac{1}{N_{i\cdot\cdot l}+\alpha_l}\right) & N_{i2\cdot\cdot} - \sum_{l=1}^{n_d} N_{i2\cdot l}^2\cdot\left(\frac{1}{N_{i\cdot\cdot l}+\alpha_l}\right) & \cdots\cdots & -\sum_{1}^{n_d} N_{i2\cdot l}\cdot N_{in_c\cdot l}\cdot\left(\frac{1}{N_{i\cdot\cdot l}+\alpha_l}\right) \\
\vdots & \ddots & & \vdots \\
-\sum_{l=1}^{n_d} N_{ij\cdot l}\cdot N_{i1\cdot l}\cdot\left(\frac{1}{N_{i\cdot\cdot l}+\alpha_l}\right) & N_{ij\cdot\cdot} - \sum_{l=1}^{n_d} N_{ij\cdot l}^2\left(\frac{1}{N_{i\cdot\cdot l}+\alpha_l}\right) & \cdots\cdots & -\sum_{1}^{n_d} N_{ij\cdot l}\cdot N_{in_c\cdot l}\left(\frac{1}{N_{i\cdot\cdot l}+\alpha_l}\right) \\
\vdots & & \ddots & \vdots \\
-\sum_{l=1}^{n_d} N_{in_c\cdot l}\cdot N_{i1\cdot l}\cdot\left(\frac{1}{N_{i\cdot\cdot l}+\alpha_l}\right) & \cdots\cdots & & N_{in_c\cdot\cdot} - \sum_{l=1}^{n_d} N_{in_c\cdot l}^2\cdot\left(\frac{1}{N_{i\cdot\cdot l}+\alpha_l}\right)
\end{bmatrix} \qquad (57)$$

Matrix $X'SZ_2$, contemporary group by sire, is given by Equation 58.

Matrix $Z_2'SZ_2$, sire, is given by Equation 59.

Matrix $X'S\underline{y}$, contemporary group right-hand side, is given by Equation 60.

Matrix $Z_2'S\underline{y}$, sire right-hand side, is given by Equation 61.

Absorption algorithm -- contemporary groups

The purpose of the within herd contemporary group absorption algorithm is to take the matrices of the dam absorption program and absorb the contemporary group effects into within herd sire effects. The PL/I program listing for this algorithm is contained in Appendix B. The algorithm accomplishes the absorption process on a herd by herd basis. The coefficients from the dam absorption program are read from the disk file and stored in appropriately defined arrays for each of the five matrices. After the arrays are filled, the direct inverse of the X'SX matrix is computed. Elements of this inverse are stored in a disk file for use in the back solution for within herd contemporary group fixed effects.

The least squares equations for sire by sire and the right-hand sides within the i-th herd are computed by Equations 62 and 63.

$$X'SZ_2 = X'Z_2 - X'Z_1(Z_1'Z_1 + G_{11}^{-1})\, Z_1'Z_2$$

$$= \begin{bmatrix} \sum_{l=1}^{n_d} N_{i111} - \sum_{l=1}^{n_d}\left[N_{i1\cdot 1}\cdot\left(\frac{1}{N_{i\cdot\cdot 1}+\alpha_1}\right)\cdot\sum_{j'=1}^{n_c} N_{ij'11}\right] & \sum_{l=1}^{n_d} N_{i121} - \sum_{l=1}^{n_d}\left[N_{i1\cdot 1}\cdot\left(\frac{1}{N_{i\cdot\cdot 1}+\alpha_1}\right)\cdot\sum_{j'=1}^{n_c} N_{ij'21}\right] & \cdots\cdots & \sum_{l=1}^{n_d} N_{i1n_s1} - \sum_{l=1}^{n_d}\left[N_{i1\cdot 1}\cdot\left(\frac{1}{N_{i\cdot\cdot 1}+\alpha_1}\right)\cdot\sum_{j'=1}^{n_c} N_{ij'n_s1}\right] \\ \sum_{l=1}^{n_d} N_{i211} - \sum_{l=1}^{n_d}\left[N_{i2\cdot 1}\cdot\left(\frac{1}{N_{i\cdot\cdot 1}+\alpha_1}\right)\cdot\sum_{j'=1}^{n_c} N_{ij'11}\right] & \sum_{l=1}^{n_d} N_{i221} - \sum_{l=1}^{n_d}\left[N_{i2\cdot 1}\cdot\left(\frac{1}{N_{i\cdot\cdot 1}+\alpha_1}\right)\cdot\sum_{j'=1}^{n_c} N_{ij'21}\right] & \cdots\cdots & \sum_{l=1}^{n_d} N_{i2n_x1} - \sum_{l=1}^{n_d}\left[N_{i2\cdot 1}\left(\frac{1}{N_{i\cdot\cdot 1}+\alpha_1}\right)\cdot\sum_{j'=1}^{n_c} N_{ij'n_s1}\right] \\ \vdots & \vdots & & \vdots \\ \sum_{l=1}^{n_d} N_{in_c11} - \sum_{l=1}^{n_d}\left[N_{in_c\cdot 1}\cdot\left(\frac{1}{N_{i\cdot\cdot 1}+\alpha_1}\right)\cdot\sum_{j'=1}^{n_c} N_{ij'11}\right] & \cdots\cdots & & \sum_{l=1}^{n_d} N_{in_cn_s1} - \sum_{l=1}^{n_d}\left[N_{in_c\cdot 1}\cdot\left(\frac{1}{N_{i\cdot\cdot 1}+\alpha_1}\right)\cdot\sum_{j'=1}^{n_c} N_{ij'n_s1}\right] \end{bmatrix}$$

(58)

$$Z_2^t S Z_2 = Z_2^t Z_2 - Z_2^t Z_1 \left(Z_1^t Z_1 + G_{11}^{-1}\right) Z_1^t Z_2$$

$$= \begin{bmatrix} N_{i\cdot 1\cdot} - \sum_{l=1}^{n_d}\left[\left(\sum_{j=1}^{n_c} N_{ijl1}\right)^2 \left(\frac{1}{N_{i\cdot\cdot 1}+\alpha_1}\right)\right] & -\sum_{l=1}^{n_d}\left[\sum_{j=1}^{n_c} N_{ijl1}\cdot\left(\frac{1}{N_{i\cdot\cdot 1}+\alpha_1}\right)\cdot\sum_{j=1}^{n_c} N_{ij21}\right] & \cdots\cdots & -\sum_{l=1}^{n_d}\left[\sum_{j=1}^{n_c} N_{ijl1}\cdot\left(\frac{1}{N_{i\cdot\cdot 1}+\alpha_1}\right)\sum_{j=1}^{n_c} N_{ijn_s 1}\right] \\ -\sum_{l=1}^{n_d}\left[\sum_{j=1}^{n_c} N_{ijl1}\left(\frac{1}{N_{i\cdot\cdot 1}+\alpha_1}\right)\cdot\sum_{j=1}^{n_c} N_{ij21}\right] & N_{i\cdot 2\cdot} - \sum_{l=1}^{n_d}\left[\left(\sum_{j=1}^{n_c} N_{ij21}\right)^2\left(\frac{1}{N_{i\cdot\cdot 1}+\alpha_1}\right)\right] & \cdots\cdots\cdots & -\sum_{l=1}^{n_d}\left[\sum_{j=1}^{n_c} N_{ij21}\left(\frac{1}{N_{i\cdot\cdot 1}+\alpha_1}\right)\sum_{j=1}^{n_c} N_{ijn_s 1}\right] \\ \vdots & & \ddots & \\ & & N_{i\cdot k\cdot} - \sum_{l=1}^{n_d}\left[\left(\sum_{j=1}^{n_c} N_{ijkl}\right)^2\left(\frac{1}{N_{i\cdot\cdot 1}+\alpha_1}\right)\right] & \vdots \\ & & & \ddots \\ -\sum_{l=1}^{n_d}\left[\sum_{j=1}^{n_c} N_{ijl1}\left(\frac{1}{N_{i\cdot\cdot 1}+\alpha_1}\right)\sum_{j=1}^{n_c} N_{ijn_s 1}\right] & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & & N_{i\cdot n_s 1} - \sum_{l=1}^{n_d}\left[\left(\sum_{j=1}^{n_c} N_{ijn_s 1}\right)^2\left(\frac{1}{N_{i\cdot\cdot 1}+\alpha_1}\right)\right] \end{bmatrix} \quad (59)$$

$$X'S\underline{y} = X'\underline{y} - X'Z_1 \left(Z_1'Z_1 + G_{11}^{-1}\right) Z_1'\underline{y}$$

$$= \begin{bmatrix} Y_{i1\cdot\cdot} - \sum_{l=1}^{n_d} \left[N_{i1\cdot l} \left(\frac{1}{N_{i\cdot\cdot l} + \alpha_l} \right) \sum_{j'=1}^{n_c} Y_{ij'\cdot l} \right] \\ Y_{i2\cdot\cdot} - \sum_{l=1}^{n_d} \left[N_{i1\cdot l} \left(\frac{1}{N_{i\cdot\cdot l} + \alpha_l} \right) \sum_{j'=1}^{n_c} Y_{ij'\cdot l} \right] \\ \vdots \\ Y_{ij\cdot\cdot} - \sum_{l=1}^{n_d} \left[N_{ij\cdot l} \left(\frac{1}{N_{i\cdot\cdot l} + \alpha_l} \right) \sum_{j'=1}^{n_c} Y_{ij'\cdot l} \right] \\ \vdots \\ Y_{in_c\cdot\cdot} - \sum_{l=1}^{n_d} \left[N_{in_c\cdot l} \left(\frac{1}{N_{i\cdot\cdot l} + \alpha_l} \right) \sum_{j'=1}^{n_c} Y_{ij'\cdot l} \right] \end{bmatrix} \qquad (60)$$

$$Z_2'S\underline{y} = Z_2'\underline{y} - Z_2'Z_1\ (Z_1'Z_1 + G_{11}^{-1})\ Z_1'y$$

$$= \begin{bmatrix} Y_{i\cdot 1\cdot} - \sum_{l=1}^{n_d}\left[\sum_{j=1}^{n_c}\left(N_{ij1l}\cdot\left(\frac{1}{N_{i\cdot\cdot l}+\alpha_l}\right)\sum_{l'=1}^{n_s} Y_{ijk'l}\right)\right] \\ Y_{i\cdot 2\cdot} - \sum_{l=1}^{n_d}\left[\sum_{j=1}^{n_c}\left(N_{ij2l}\cdot\left(\frac{1}{N_{i\cdot\cdot l}+\alpha_l}\right)\sum_{k'=1}^{n_s} Y_{ijk'l}\right)\right] \\ \vdots \\ Y_{i\cdot k\cdot} - \sum_{l=1}^{n_d}\left[\sum_{j=1}^{n_c}\left(N_{ijkl}\cdot\left(\frac{1}{N_{i\cdot\cdot l}+\alpha_l}\right)\sum_{k'=1}^{n_s} Y_{ijk'l}\right)\right] \\ \vdots \\ Y_{i\cdot n_s\cdot} - \sum_{l=1}^{n_d}\left[\sum_{j=1}^{n_c}\left(N_{ijn_sl}\cdot\left(\frac{1}{N_{i\cdot\cdot l}+\alpha_l}\right)\sum_{k'=1}^{n_s} Y_{ijk'l}\right)\right] \end{bmatrix} \qquad (61)$$

$$Z_2'S^*Z_2 = Z_2'SZ_2 - Z_2'SX(X'SX)^{-1} X'SZ_2 \quad (62)$$

$$Z_2'S^*\underline{y} = Z_2'S\underline{y} - Z_2'SX(X'SX)^{-1} X'S\underline{y} \quad (63)$$

where

$$S^* = S - SX(X'SX)^{-1} X'.$$

Each coefficient for $Z_2'S^*Z_2$ and $Z_2'S^*\underline{y}$ is written to a disk file for sorting and summing after all herds have been processed. Each coefficient row and column is labeled by the appropriate sire registration number for $Z_2'S^*Z_2$, and by the sire registration number and 99999991 for $Z_2'S^*\underline{y}$ matrix. After all herds are processed, the disk file containing coefficients from each herd is sorted by row and column and then appropriately summed within identical row and column indices to form the final mixed model sire evaluation equations.

Positive off-diagonals in $Z_2'S^*Z_2$ are possible by including dam effects in the model. This is contrary to what occurs in the current sire evaluation model that does not include dam effects. All off-diagonals are negative in the current sire evaluation model. In both cases, however, off-diagonals sum to the diagonal element. The positive off-diagonals occur when negative terms appear in the $X'SZ_2$ matrix. The terms in this matrix are all positive for the current sire evaluation methodology. The first term in the $X'SZ_2$ matrix is the number of progeny records the k-th sire has in the j-th contemporary group. This first term is reduced as a function of the number of progeny each dam, also with progeny in the j-th contemporary group, has in other contemporary groups. The reduction is also influenced by the number

of times the k-th sire is repeat mated to these dams. The higher the repeat matings, the more negative the $X'SZ_2$ term becomes.

The remainder of the national sire evaluation is done with operational sire evaluation programs at Iowa State University. These programs perform the task of building the inverse of Wright's (1922) numerator relationship matrix for sires and maternal grandsires to add to the sire equations by the method of Henderson (1975c), building sire birth year groups, and iterating the final set of equations to determine birth year group fixed effect estimates and sire effect predictors.

Back solutions -- contemporary group effects

The purpose of this program is to retrieve within herd contemporary group matrices ($X'SX^{-1}$, $X'SZ_2$, and $X'S\underline{y}$) and solve for the effects using the most recently determined sire EPDs for direct WWT. The general expression for determining these fixed effect estimates is given by the following:

$$\hat{\underline{c}}_i = [X'SX]_i^{-1} \left[(X'Sy)_i - (X'SZ_2)_i \hat{\underline{u}}_2^*\right] \tag{64}$$

where

$\hat{\underline{c}}_i$ = an unknown vector of contemporary group fixed effects for the i-th herd, and

$\hat{\underline{u}}_2^*$ = a known vector of sire EPDs for direct WWT.

The procedure works on a herd by herd basis and starts by building arrays for each of the three matrices as their coefficients are retrieved from the disk storage file. The effects are determined directly by

Equation 64 by selecting only those EPDs for sires with progeny in the given herd. The solutions are stored for future access as required in back solutions for dam effects and in estimation of environmental trends.

Back solutions -- dam composite effects

The purpose of the algorithm is to compute the composite dam effect. The effect includes both the genetic direct effect and maternal effect for weaning weight. The algorithm is contained in Appendix D.

The algorithm is run following the back solution algorithm for contemporary group effects. The process is done on a herd by herd basis. Within herd contemporary group effect solutions are retrieved and temporarily stored in an array indexed by the contemporary group identification. All sire EPDs are retrieved and stored in a vector for use in determining dam effects across all herds. Following storage, the original performance data set is read in a sequential manner on a dam by dam basis. Each progeny record for a given dam is adjusted for the contemporary group estimate and sire's EPD for direct WWT. The adjusted records are summed for each dam and regressed by the total number of records the dam has and by the error variance to dam variance ratio. The general solution for all dam composite effects within the i-th herd is given by the following expression:

$$\hat{\underline{u}}_1 = \left[(Z_1'Z_1)_i + \sigma_e^2/\sigma_d^2 \cdot I\right]^{-1} \cdot$$

$$\left[(Z_1'\underline{y})_i - (Z_1'X)_i\ \hat{\underline{c}}_i - (Z_1'Z_2)\ \hat{\underline{u}}_2^*\right] \qquad (65)$$

where

$\hat{\underline{u}}_1$ = an unknown vector of dam composite effects within the i-th herd for direct and maternal WWT,

$\hat{\underline{c}}_i$ = a known vector of contemporary group fixed effects within the i-th herd, and

$\hat{\underline{u}}^*_2$ = a known vector of sire EPDs for direct WWT.

The actual computation to determine dam composite effects is done on a dam by dam basis, and for the 1-th dam is given by:

$$\hat{d}_{il} = (n_{i\cdot\cdot 1} + \alpha)^{-1} \left(Y_{i\cdot\cdot 1} - \sum_{j=1}^{n_c} n_{ij\cdot 1}\hat{c}_{ij} - \sum_{k=1}^{n_s} n_{i\cdot k1}\hat{s}^*_k\right) \qquad (66)$$

where

$\hat{d}_{il}$ = a composite effect for direct and maternal WWT for the 1-th dam in the i-th herd,

$Y_{i\cdot\cdot 1}$ = the sum of all progeny performance records for the 1-th dam,

$\hat{c}_{ij}$ = known fixed effect for the j-th contemporary group of the i-th herd,

$\hat{s}^*_k$ = known EPD for direct WWT for the k-th sire, and

$\alpha = \sigma^2_e/\sigma^2_d$.

Additional information can be computed and stored for future reference since all effects are known when this algorithm is run. The information includes:

1) Total sums of squares (SS) of the observations,

2) The sum of squares (SS) for each contribution to the model, including sire birth year SS, contemporary group SS, sire SS, and dam SS,

3) REML estimate of model error variance using the total SS less model SS divided by the number of observations as adjusted for the number of birth year groups and contemporary groups, and

4) Within herd phenotypic, environmental, and genetic contributions to each progeny birth year group (the environmental and genetic contributions are weighted by the number of progeny in each contemporary group and by the number of progeny each sire and dam has, respectively).

The computational algorithms and procedures are based upon assumptions that are made in the development of the evaluation equations. The algorithms and procedures must include appropriate estimates of the parameters of the population or data set being evaluated. Variance component estimation of the parameters is an integral part of evaluation model development. The next section of the chapter discusses variance component estimation.

Variance Component Estimation

Mixed model sire and within herd evaluations assume that random effect variance and covariance parameters are known. True genetic parameter values are never known, but good estimates may be available from prior data. If there are no prior estimates, or if available estimates are considered invalid for the particular data set being used, a method of parameter estimation is needed. The method of parameter estimation chosen must be computationally feasible and must supply parameter estimates that are unbiased and have small sampling variance.

Prior estimates do not exist for the data used and the evaluation models proposed in this thesis. Determining the estimates will

constitute a major research effort. This section discusses approaches to the problem of determining variance and covariance parameter estimates for the following two mixed models:

1) The direct WWT sire evaluation model with dam effects included, and

2) The maternal WWT sire evaluation model.

Direct WWT sire evaluation model

The direct WWT sire evaluation model uses a mixed model to determine best linear unbiased predictors for sire EPDs. Several methods available to estimate variance components associated with a model of this type include: Henderson's Method 3 (1953), Hartley and Rao's maximum likelihood (ML) (1967), Patterson and Thompson's restricted maximum likelihood (REML) (1971), Rao's minimum norm quadratic unbiased estimation (MINQUE) (1971), and approximating the REML approach (Harville, 1977; and Henderson, 1980). Computational difficulty varies with each method and depends upon the number of levels of random effects in the evaluation data set.

The number of levels of random effects in many animal breeding evaluation models is extremely large. For that reason, many methods of variance component estimation are not practical due to the computational difficulty. The methods, with the exception of the approximate REML approach, require a complete inverse of the coefficient matrix. The approximate REML procedure used in this thesis is based upon Henderson (1980) and is referred to as Henderson's new method.

The within herd dam mixed model equations range in matrix order from 35 to 1,459. There are 20 herds in the validation data base with

an average number of 520 dams per herd. Each dam represents one level of a random effect. The sire mixed model equations are of matrix order 982. It would be possible to restrict the number of herds used in parameter estimation. However, the size of the coefficient matrix after absorption of fixed effects would still be excessively large for four of the methods listed above. Reduction of the data size to be compatible with the methods is not an attractive alternative because poor parameter estimates are likely.

Henderson's new method, in contrast to Method 3, ML, REML, and MINQUE, requires no matrix inverse if X'X is diagonal and the inverse of X'X only if it is not diagonal (Henderson, 1980). Therefore, Henderson's new method appears feasible with the data base being used in this paper for unified sire and within herd evaluation models.

Henderson's new method is similar to Method 3 in that quadratic forms are equated to their expectations to form a set of equations to be solved simultaneously for the variance component estimates. The quadratic forms in the new method are computed from approximations to best linear unbiased predictors and are equated to their expectations. As in Method 3, more than one set of quadratic forms exists. Hudson and Van Vleck (1982) considered u'u and u'r quadratic forms to compare numerical estimates and computing times of Henderson's new method to Method 3.

Henderson's method can be used in an iterative way similar to iteration in REML or other iterative methods. However, iteration in the new method is not required. Hudson and Van Vleck (1982) pointed

out that properties of the variance components obtained iteratively are unknown. Iterative estimates may be biased in contrast to first round estimates that are unbiased. The researchers found that convergence occurred rapidly when using iteration.

Numerical variance components obtained by Hudson and Van Vleck (1982) were similar to Method 3 estimates for most traits and data sets considered. Relative computing time for Method 3 and Henderson's new method depended on the number of equations and right-hand sides. If sire genetic groups had not been included in addition to herd-year-season fixed effects, the new method would always have been computationally faster. The group equations were formed from sire equations after absorption of the herd-year-season fixed effects. Absorption of group equations was a costly process.

Dempfle et al. (1983) compared the efficiency of Henderson's new method to MINQUE. For the comparison, inefficiency was defined as:

$$\text{Inefficiency} = \frac{\text{Var}\ (\hat{\sigma}^2 \mid \text{for given data set and procedure})}{\text{Var}\ (\hat{\sigma}^2 \mid \text{for all data and best procedure})}.$$

Dempfle et al. found that by using an appropriate value of $\hat{h}^2$ (estimated heritability), such that $|h^2 - \hat{h}^2|$ was small, the new method's estimator was inferior to the MINQUE estimator. However, the new method always retained a high efficiency. The efficiency was highest for very small h^2. In their data set, the inefficiency for Henderson's new method was 1.013 for $h^2 = \hat{h}^2 = .01$ and 1.151 for $h^2 = \hat{h}^2 = 1$. The researchers concluded that if MINQUE is impossible or very difficult to compute, Henderson's new method seems to be a useful alternative.

Application of Henderson's new method to the direct WWT sire evaluation model that includes dam effects is presented in this thesis. The development is patterned after an example given by Schaeffer (1983), but assumes a different variance-covariance structure between the random effects. Henderson's (1980) presentation of the new method as well as Schaeffer's example assume that sire random effects are uncorrelated. The variance component estimation procedure developed in this thesis assumes relationship ties among sires. This assumption is consistent with the sire evaluation mixed model variance-covariance matrix.

Mixed model equations associated with sire and dam evaluation model are as follows:

$$\begin{bmatrix} X'X & X'Z_1 & X'Z_2 \\ Z_1'X & \begin{pmatrix} Z_1'Z_1 & Z_1'Z_2 \\ Z_2'Z_1 & Z_2'Z_2 \end{pmatrix} + G^{-1} \end{bmatrix} \begin{bmatrix} \tilde{c} \\ \tilde{u}_1 \\ \tilde{u}_2 \end{bmatrix} = \begin{bmatrix} X'y \\ Z_1'y \\ Z_2'y \end{bmatrix} \qquad (67)$$

where

$$G^{-1} = \sigma_e^2 \left[\begin{array}{c|c} \sigma_d^2 \cdot I & 0 \\ \hline 0 & \sigma_s^2 \cdot A \end{array}\right]^{-1}.$$

The following assumptions apply to this model:

$$E(u_1) = 0 \qquad \text{Var } (u_1) = I \cdot \sigma_d^2$$

$$E(u_2) = 0 \qquad \text{Var } (u_2) = A \cdot \sigma_s^2$$

$$E(e) = 0 \qquad \text{Var } (e) = I \cdot \sigma_e^2$$

$$E(y) = Xc \qquad \text{Var } (y) = Z_1 Z_1' \sigma_d^2 + Z_2 A Z_2' \sigma_s^2 + I \cdot \sigma_e^2 = V.$$

Henderson's new method and the estimation procedure used in this thesis to compute quadratic forms and their expectations are outlined in the following steps:

Step 1: Assume values for σ_e^2/σ_d^2 and σ_e^2/σ_s^2 and solve the mixed model equations for $\underline{\hat{c}}$, $\underline{\hat{u}}_1$, and $\underline{\hat{u}}_2$.

Step 2: Compute a REML estimate for σ_e^2 as given by the following equation:

$$\sigma_e^2 \cong \frac{\underline{y}'\underline{y} - \underline{\hat{c}}'X'\underline{y} - \underline{\hat{u}}'Z\underline{y}}{N - \text{rank}(X)} \tag{68}$$

where

$\underline{y}'\underline{y}$ = total sums of squares of observations,

N = total number of observations,

$\underline{\hat{c}}'X'\underline{y}+\underline{\hat{u}}'Z'\underline{y}$ = model sums of squares, and

$\text{rank}(X)$ = number of contemporary groups.

Step 3: Set up the least squares equations:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z \end{bmatrix} \begin{bmatrix} \underline{\tilde{c}} \\ \underline{\tilde{u}} \end{bmatrix} = \begin{bmatrix} \underline{r}_1 \\ \underline{r}_2 \end{bmatrix} \tag{69}$$

where

$$Z = (Z_1' \vdots Z_2'),$$

$$\underline{\tilde{u}} = (\underline{\tilde{u}}_1' \vdots \underline{\tilde{u}}_2').$$

Step 4: Absorb the fixed effects into the random effects such that:

$$\begin{bmatrix} Z'MZ \end{bmatrix} \begin{bmatrix} \underline{\tilde{u}} \end{bmatrix} = \begin{bmatrix} Z'M\underline{y} \end{bmatrix} \tag{70}$$

where

$$M = I - X(X'X)^{-1}X'.$$

Step 5: Select an approximation to the best linear unbiased predictor $\underline{u}_i$ to be:

$$\hat{\underline{u}}_i = D_i^{-1} Z_i'M\underline{y} \tag{71}$$

where

$$D_1 = \text{Diagonal } (Z_1'MZ_1 + \sigma_e^2/\sigma_d^2 \cdot I),$$

$$D_2 = \text{Diagonal } (Z_2'MZ_2 + \sigma_e^2/\sigma_s^2 \cdot A^{-1}).$$

Step 6: Using the vector of approximations found in Step 5, compute the quadratic forms $\hat{\underline{u}}_i'\hat{\underline{u}}_i$ for i=1 and $\hat{\underline{u}}_i'A^{-1}\hat{\underline{u}}_i$ for i=2. This would be the dam composite effect EPDs and sire direct WWT EPDs as approximated by Equation 71.

Step 7: Determine the expectations for the quadratic forms found in Step 6. The development of the expectations for u'u follows:

$$\begin{aligned}\hat{\underline{u}}_i'\,\hat{\underline{u}}_i &= \underline{y}'MZ_i(D_i^{-1})D_i^{-1}Z_i'M\underline{y} \\ &= \underline{r}_i'D_i^{-2}\underline{r}_i \\ &= \underline{r}'Q_i\underline{r} \end{aligned} \qquad (72)$$

where

$$\underline{r}' = (\underline{r}_1' \vdots \underline{r}_2'),$$

$$Q_1 = \left[\begin{array}{c|c} D_1^{-2} & \phi \\ \hline \phi & \phi \end{array}\right], \qquad Q_2 = \left[\begin{array}{c|c} \phi & \phi \\ \hline \phi & D_2^{-2} \end{array}\right].$$

Now the expectation is given by:

$$E(\underline{r}'Q_i\underline{r}) = E(\underline{r}')Q_iE(\underline{r}) + \text{trace } (Q_i\text{Var}(\underline{r}))$$

where

$$\begin{aligned} E(\underline{r}) &= E(Z'M\underline{y}) \\ &= Z'ME(\underline{y}) \\ &= Z'MX\underline{c} \\ &= Z'(I - X(X'X)^{-1}X')X\underline{c} \\ &= Z'\phi. \end{aligned} \qquad (73)$$

The Var($\underline{r}$) is given by:

$$\begin{aligned}\text{Var}(\underline{r}) &= \text{Var}(Z'M\underline{y}) \\ &= Z'MVMZ \\ &= Z'M(Z_1Z_1'\sigma_d^2 + Z_2AZ_2'\sigma_s^2 + \sigma_e^2)MZ \\ &= \sum_{i=0}^{2} Z'MZ_iG_iZ_i'MZ \\ &= \sum_{i=0}^{2} C_i\sigma_i^2 \end{aligned} \tag{74}$$

where

$$C_i = Z'MZ_iG_iZ_i'MZ\ \sigma_i^2,$$

G_0, G_1, G_2 = I, I, and A, respectively, and

$\sigma_0^2, \sigma_1^2, \sigma_2^2 = \sigma_e^2, \sigma_d^2, \sigma_s^2$, respectively.

The expectation is given as:

$$\begin{aligned}E(\underline{r}'Q_i\underline{r}) &= 0 + \text{trace}(Q_i \sum_{i=0}^{2} C_i\sigma_i^2) \\ &= \sum_{j=0}^{2} \text{trace}(Q_iC_j\sigma_j^2). \end{aligned} \tag{75}$$

Step 8: Solve the following set of equations simultaneously for σ_d^2 and σ_s^2:

$$\begin{bmatrix} \text{tr}(Q_1c_1) & \text{tr}(Q_1c_2) & \text{tr}(Q_1c_0) \\ \text{tr}(Q_2c_1) & \text{tr}(Q_2c_2) & \text{tr}(Q_2c_0) \end{bmatrix} \begin{bmatrix} \hat{\sigma}_d^2 \\ \hat{\sigma}_s^2 \\ \hat{\sigma}_e^2 \end{bmatrix} = \begin{bmatrix} \hat{\underline{u}}_1'\hat{\underline{u}}_1 \\ \hat{\underline{u}}_2'A^{-1}\hat{\underline{u}}_2 \end{bmatrix}. \tag{76}$$

Step 9: Use the estimates found in Step 8 back in Step 1 to start another round of the iteration process. The estimates for each round become the priors for each subsequent round of iteration. Continue the iteration process until the estimates converge.

Aside from absorbing fixed effects into the random effects, the major task associated with Henderson's new method lies in computing each trace$(Q_i C_i)$. Starting with the quantity $Q_1 C_1$, several of the traces are developed as follows:

$$Q = \begin{bmatrix} (n'_{11}+\sigma_e^2/\sigma_d^2)^{-2} & & \phi & \phi \\ & (n'_{12}+\sigma_e^2/\sigma_d^2)^{-2} & & \\ & & \ddots & \\ \phi & & (n'_{1n_d}+\sigma_e^2/\sigma_d^2)^{-2} & \\ \hline & \phi & & \phi \end{bmatrix}$$

$$= \begin{bmatrix} q_{11} & & & \phi \\ & q_{12} & & \\ & & \ddots & \\ & & q_{1n_d} & \\ \hline & \phi & 0 \quad \phi & \\ & & \ddots & \\ & & \phi \quad 0 & \end{bmatrix}$$

$$= \begin{bmatrix} D_1 & 0 \\ \hline 0 & 0 \end{bmatrix}. \qquad (77)$$

$$C_1 = Z'MZ_1IZ_1'MZ$$

$$\begin{bmatrix} Z_1'MZ_1Z_1'MZ_1 & Z_1'MZ_1Z_1'MZ_2 \\ Z_2'MZ_1Z_1'MZ_1 & Z_2'MZ_1Z_1'MZ_2 \end{bmatrix}$$

$$= \begin{bmatrix} C^{11} & C^{12} \\ C_1^{21} & C_1^{22} \end{bmatrix}. \tag{78}$$

Therefore,

$$Q_1 C_1 = \left[\begin{array}{c|c} D_1 & 0 \\ \hline 0 & 0 \end{array}\right] \left[\begin{array}{c|c} C_1^{11} & C_1^{12} \\ \hline C_1^{21} & C_1^{22} \end{array}\right]$$

$$= \left[\begin{array}{c|c} D_1\, C_1^{11} & D_1\, C_1^{12} \\ \hline 0 & 0 \end{array}\right]. \tag{79}$$

Now the trace$(Q_1 C_1)$ is given by:

$$\text{trace}(Q_1 C_1) = \text{trace} \left[\begin{array}{c|c} D_1\, C_1^{11} & D_1\, C_1^{12} \\ \hline 0 & 0 \end{array}\right]$$

$$= \Sigma (\text{diagonal elements of } D_1\, C_1^{11}). \tag{80}$$

Since D_1 is diagonal, only the diagonal elements of C_1^{11} are required to compute this trace. Therefore,

$$\text{trace}(Q_1 c_1) = \sum_{k=1}^{n_d} q_{kk}\, C_{1(kk)}^{11}$$

$$= \sum_{k=1}^{n_d} (n'_{1k} + \sigma_e^2/\sigma_d^2)^{-2} \text{Diag}(Z_1' M Z_1)^2 \tag{81}$$

where n'_{1k} is equal to the diagonal element of the dam equations after absorption of contemporary groups. Similarly, it can be shown that

$$\text{trace}(Q_1 C_2) = \sum_{k=1}^{n_d} q_{kk}\, C_{2(kk)}^{11}$$

$$= \sum_{k=1}^{n_d} (n'_{1k} + \sigma_e^2/\sigma_d^2)^{-2} \text{Diag}(Z_1' M Z_2 A Z_2' M Z_1). \tag{82}$$

The trace($Q_1 C_2$) is of considerable more difficulty to compute than the trace($Q_1 C_1$) because of the relationship matrix, A. The trace($Q_i C_0$) is given by:

$$\text{trace}(Q_1 C_0) = \sum_{k=1}^{n_d} q_{kk} \; C_{0\,(kk)}^{11} \tag{83}$$

where

$$C_0 = \begin{bmatrix} Z_1'MZ_1 & Z_1'MZ_2 \\ Z_2'MZ_1 & Z_2'MZ_2 \end{bmatrix}.$$

The traces of ($Q_2 C_i$), i = 0, 1, and 2, are determined in the same manner as above. After all the traces are computed, Equation 76 can be solved for the first round of estimates for σ_d^2 and σ_s^2. The new estimates are used to begin another round of estimation.

Maternal WWT sire evaluation model

The maternal WWT sire evaluation model uses mixed model equations to determine sire EPDs for maternal WWT. Fixed effects are birth years of sires in the model. The random effects include maternal grandsire (MGS) direct and maternal effects and a residual error effect. Direct and maternal random effects are assumed correlated, and, therefore, both variances and covariances are required for the mixed model. The residual error effects are assumed uncorrelated. The model equation is given by the following:

$$\underline{y} = X_{\underline{g}} + Z_1 \underline{u}_1 + Z_2 \underline{u}_2 + \underline{e} \tag{84}$$

where

$\underline{y}$ = a known vector of daughter's progeny WWT records adjusted for contemporary group fixed effects and sire of progeny direct effect,

X = a known incidence defining the birth year of each MGS in the model,

$\underline{g}$ = an unknown vector of MGS birth year fixed effects,

Z_1, Z_2 = known incidence matrices relating MGSs to records of daughters with Z_1 associated with the direct effect for WWT and containing 0's and .5's, Z_2 is associated with the maternal effect for WWT and contains 0's and 1's,

$\underline{u}_1, \underline{u}_2$ = vectors of random effects for direct and maternal WWT, respectively. The vector $\underline{u}_1$ is known, and the vector $\underline{u}_2$ is unknown. The effects are assumed bivariate normal with $E(\underline{u}_i)=0$, and

$\underline{e}$ = a vector of residual random effects, potentially correlated due to record adjustments, but assumed NID $(0,\sigma_e^2)$.

The mixed model equations for the maternal WWT evaluation model are given by the following:

$$\begin{bmatrix} X'X & X'Z_1 & X'Z_2 \\ Z_1'X & \begin{pmatrix} Z_1'Z_1 & Z_1'Z_2 \\ Z'Z_1 & Z_2'Z_2 \end{pmatrix} + G^{-1} \\ Z_2'X & \end{bmatrix} \begin{bmatrix} \tilde{g} \\ \tilde{u}_1 \\ \tilde{u}_2 \end{bmatrix} = \begin{bmatrix} X'\underline{y} \\ Z_1'\underline{y} \\ Z_2'\underline{y} \end{bmatrix} \qquad (85)$$

where

$$G^{-1} = \sigma_e^2 \left[\begin{array}{c|c} \sigma_{s_D}^2 \cdot I & \sigma_{s_D s_M} \cdot I \\ \hline \sigma_{s_D s_M} I & \sigma_{s_M}^2 \cdot I \end{array}\right]^{-1}.$$

There are two distinguishing factors associated with estimating the required variance components for the mixed model equations of Equation 85. First, there is only one observation for the two traits

of interest, a calf's WWT record. Each adjusted WWT record is a measure of the genetic direct effect and the maternal effect combined. Secondly, the variance for the direct effect can be determined independently of the variance for the maternal effect and the covariance. The method of obtaining an estimate for variance of the sire's direct effect for WWT is given in the preceding part of this section.

The following paragraphs outline a possible method of determining the variance of sire maternal effect ($\sigma^2_{s_D}$) and the covariance between sire direct effect and sire maternal effect ($\sigma_{s_D s_M}$). The development presented in this thesis is an interpretation of variance component estimation given by Searle (1979) and Schaeffer (1983). The method follows a general strategy used for estimation of variance components: (1) define quadratic forms, (2) calculate the expectations of the quadratic forms, and (3) equate the quadratics to their expectations to solve for the variance components.

Quadratic forms A list of quadratic forms which have expectations involving the variance components of interest are as follows:

$$u_1^{*\prime}u_1^*$$

$$u_2^{*\prime}u_2^*$$

$$u_1^{*\prime}u_2^*$$

where u_i^* is a BLUP found from Equation 85 using an a priori value for G^{-1} equal to G^{*-1}.

Let the generalized inverse of the coefficient matrix of Equation 85 be written as follows:

$$\begin{bmatrix} c_{00} & c_{01} & c_{02} \\ c_{10} & c_{11} & c_{12} \\ c_{20} & c_{21} & c_{22} \end{bmatrix},$$

then BLUP solutions using G^{*-1} can be determined as follows:

$$\begin{bmatrix} \hat{\underline{g}}^* \\ \hat{\underline{u}}_1^* \\ \hat{\underline{u}}_2^* \end{bmatrix} = \begin{bmatrix} c_{00} & c_{01} & c_{02} \\ c_{10} & c_{11} & c_{12} \\ c_{20} & c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} X'\underline{y} \\ Z_1'\underline{y} \\ Z_2'\underline{y} \end{bmatrix}$$

$$= \begin{bmatrix} c_0\underline{r} \\ c_1\underline{r} \\ c_2\underline{r} \end{bmatrix}. \tag{86}$$

Expectations The expectations for the quadratic forms can be written as:

$$E(\underline{u}_1^{*\prime}\underline{u}_1^*) = E(\underline{r}'c_1'c_1\underline{r}')$$

$$E(\underline{u}_2^{*\prime}\underline{u}_2^*) = E(\underline{r}'c_2'c_2\underline{r})$$

$$E(\underline{u}_1^{*\prime}\underline{u}_2^*) = E(\underline{r}'c_1'c_2\underline{r}). \tag{87}$$

The expectation of the last quadratic given in Equation 87 will be addressed first. The expectation is given as follows:

$$E(\underline{r}'c_1'c_2\underline{r}) = E(\underline{r}')c_1'c_2E(\underline{r}) + \text{trace}(c_1'c_2V(\underline{r})) \tag{88}$$

where $V(\underline{r})$ is equal to the variance of $\underline{r}$. The variance components to be estimated have the property of being "translation invariant" if

$c_1'c_2X=\phi$. That is, the variance component estimates are not affected by fixed effects and the expectation given in Equation 88 can be reduced to:

$$E(\underline{r}'c_1'c_2\underline{r}) = \text{trace}(c_1'c_2V(\underline{r})). \tag{89}$$

The variance of $\underline{r}$ is given as follows:

$$V(r) = \begin{bmatrix} X' \\ Z_1' \\ Z_2' \end{bmatrix} V \begin{bmatrix} X & Z_1 & Z_2 \end{bmatrix}$$

$$= W'VW$$

where V is equal to the variance of $\underline{y}$. The variance matrix V is given as follows:

$$V = \begin{bmatrix} X & Z_1 & Z_2 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & \sigma_1^2 \cdot I & \sigma_{12} \cdot I \\ 0 & \sigma_{12} \cdot I & \sigma_2^2 \cdot I \end{bmatrix} \begin{bmatrix} X \\ Z_1 \\ Z_2 \end{bmatrix} + \sigma_0^2 \cdot I$$

$$= Z_1Z_1'\sigma_1^2 + (Z_1Z_2' + Z_2Z_1')\sigma_{12} + Z_2Z_2'\sigma_2^2 + \sigma_0^2 I$$

$$= V_1 + V_3 + V_2 + V_0$$

where

σ_1^2 = the variance of sire direct effect for WWT,

σ_2^2 = the variance of sire maternal effect for WWT,

σ_{12} = the covariance of sire direct effect and maternal effects for WWT, and

σ_0^2 = the model residual error variance.

The resulting expectation of $u_1^{*\prime}u_2^*$ can be written as follows:

$$E(u_1^{*\prime}u_2^*) = \text{trace}(c_1'c_2 \sum_{j=0}^{3} V_i)$$

$$= \sum_{i=0}^{3} \text{trace}(c_1'c_2V_i). \tag{90}$$

Similarly, the other two quadratic form expectations can be written as follows:

$$E(u_1^{*\prime}u_1^*) = \sum_{i=0}^{3} \text{trace}(c_1'c_1V_i) \tag{91}$$

$$E(u_2^{*\prime}u_2^*) = \sum_{i=0}^{3} \text{trace}(c_2'c_2V_i). \tag{92}$$

The traces given by Equations 90, 91, and 92 would require several large matrix product calculations to be computed. Some simplification can be obtained using the relationships given as follows:

$$V(r) = W'VW$$

$$= W'Z_1Z_1'W\sigma_1^2 + W'Z_1Z_2'W\sigma_{12} + W'Z_2Z_1'W\sigma_{12} + W'Z_2Z_2'W\sigma_2^2$$

$$+ W'W\sigma_0^2$$

$$\therefore E(u_1^{*\prime}u_2^*) = \text{trace}(c_1'c_2W'VW)$$

$$= \text{trace}(c_1'c_2W'Z_1Z_1'W\sigma_1^2)$$

$$+ \text{trace}(c_1'c_2W'Z_1Z_2'W\sigma_{12})$$

$$+ \text{trace}(c_1'c_2W'Z_2Z_1'W\sigma_{12})$$

$$+ \text{trace}(c_1'c_2W'Z_2Z_2'W\sigma_2^2)$$

$$+ \text{trace}(c_1'c_2W'W\sigma_0^2). \tag{93}$$

The following relationship from Henderson (1975f) is useful in the simplification process:

$$\begin{bmatrix} c_{00} & c_{01} & c_{02} \\ c_{10} & c_{11} & c_{12} \\ c_{20} & c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} X'X & X'Z_1 & X'Z_2 \\ Z_1'X & Z_1'Z_1 & Z_1'Z_2 \\ Z_2'X & Z_2'Z_1 & Z_2'Z_2 \end{bmatrix} =$$

$$\begin{bmatrix} I_0 & -c_{01}G^{-1} & -c_{02}G^{-1} \\ \phi & I_1 - c_{11}G^{-1} & -c_{12}\cdot G^{-1} \\ \phi & -c_{21}\cdot G^{-1} & I_2 - c_{22}G^{-1} \end{bmatrix} \tag{94}$$

where the elements in this relationship can be used in the traces given by Equation 93. From the result that trace (ABC) = trace (CAB), the following equation can be written:

$$\text{trace}(c_1'c_2W'Z_1Z_1'W)\sigma_1^2 = \text{trace}((Z_1'Wc_1')(c_2W'Z_1))\sigma_1^2. \tag{95}$$

The matrix product $(c_2W'Z_1)$ can be written with partitioned matrices from Equation 94 as follows:

$$(c_2W'Z_1) = (c_{20}c_{21}c_{22}) \begin{bmatrix} X' \\ Z_1' \\ Z_2' \end{bmatrix} (Z_1)$$

$$= -c_{21}\cdot G^{-1}.$$

Similarly,

$$(c_1W'Z_1) = I - c_{11}G^{-1}$$

so that

$$\text{trace } ((Z_1'Wc_1')(c_2W'Z_1))\sigma_1^2 = \text{trace } ((I-c_{11}G^{-1})'(-c_{21}\cdot G^{-1}))\sigma_1^2. \quad (96)$$

Each of the trace operations in Equation 93 can be reexpressed with terms from Equation 94. The expected values of the quadratics $u_1^{*'}u_1^*$ and $u_2^{*'}u_2^*$ can also be reexpressed with terms from Equation 94.

Equating quadratic and expectations The last step in determining the variance component estimates is to equate the quadratics of BLUP solutions using the apriori G^{*-1} to their expectations and solve for $\hat{\sigma}_2^2$, $\hat{\sigma}_{12}$, and $\hat{\sigma}_0^2$. The equations to be solved are as follows:

$$\begin{aligned} u_1^{*'}u_1^* &= \text{trace } (c_1'c_1W'VW) \\ u_1^{*'}u_2^* &= \text{trace } (c_1'c_2W'VW) \\ u_2^{*'}u_2^* &= \text{trace } (c_2'c_2W'VW). \end{aligned} \quad (97)$$

Recall that $\hat{\sigma}_1^2$ will be known and need not be estimated again. An estimate for residual error variance (σ_0^2) can be obtained from the following:

$$\hat{\sigma}_0^2 = (\underline{y}'\underline{y} - \underline{g}^{*'}X'\underline{y} - \underline{u}^{*'}Z'\underline{y})/(N-\text{rank}(X)). \quad (98)$$

If σ_0^2 is estimated using Equation 98, then only two equations from Equation 97 would be needed to solve for σ_{12} and σ_2^2.

Considerable research effort is still needed to insure the appropriateness of the above procedure to obtain variance component estimates for the sire maternal WWT evaluation model or whether some

other procedure is warranted. The full model without adjustments for fixed effects and sire effects may be more appropriate to reduce the likelihood of a correlated residual error structure. The next chapter presents a summary and discussion of some of the thesis results.

RESULTS AND DISCUSSION

This section begins with a discussion of computational aspects for the proposed evaluation methodology of unifying national beef sire evaluation and within herd evaluations. Computational time projections are made for portions of an operational evaluation based upon the times observed with evaluating the validation data base. The following section of this chapter presents a comparison of sire EPDs for direct WWT obtained from the proposed evaluation model using the validation herds. Direct WWT EPDs are obtained for all sires used in the validation herds and are compared with the EPDs determined in the 1983 Angus sire evaluation. The last section addresses preliminary results obtained in assessing the merit of the method for calculating sire prediction accuracies in the 1983 Angus and Hereford sire evaluations.

Computational Aspects

The computational aspects discussed in the chapter refer to four major computer algorithms developed as a part of the national sire and within herd evaluation unification. Time and region projections are derived from an analysis of 20 validation Angus herds. Many of the herds included have a long history of performance testing with an above average number of performance records. Therefore, projections of computational times should be conservative since many Angus herds are considerably smaller and contain fewer performance records.

Table 5 summarizes the computational times associated with the four major programs. The times are given in terms of total time for

Table 5. Computer time requirements for the proposed sire/dam evaluation model

Program	Execution times (CPU min.)[a]			
	Total	Average per herd	Average per 100,000 records	Region (K)
Absorption:				
Dams	0.412	0.0206	1.355	112
Contemporary groups	0.648	0.0324	2.131	988
Sorting and summing	0.839	0.0419	2.761	256
Total	1.899	0.0949	6.247	

[a]Based on 30,411 WWT records.

the 20 validation herds, in terms of an average time per herd, and in terms of an average time per 100,000 performance records. The programs are designed to be used in conjunction with programs currently being used at Iowa State University to conduct a national beef sire evaluation that includes dam effects. The current contemporary group absorption program would be replaced with the new contemporary group absorption program.

Table 6 lists computational times associated with the Iowa State University national beef sire evaluation programs that were obtained during the 1983 Angus Sire Evaluation for direct WWT. Starting with more than 805,000 WWT records, 354,746 remain after editing for bad records and removing contemporary groups with only one sire represented (Berger et al., 1983).

Table 6. Computer time requirements for current sire evaluation

Program	Execution time (CPU min.) Total	Average per 100,000 records	Region (K)
Data editing	17.82[a]	2.214	150
Contemporary group absorption	2.82[b]	0.355	116
Adding relationships	2.19[c]	0.617	576
Equation iterations	47.22[c]	13.339	496
Miscellaneous	2.00	0.250	256
Total	72.05	16.775	

[a]Based upon 805,000 unedited field records.

[b]Based upon 795,134 edited WWT records.

[c]Based upon 354,746 weaning weight records and 28,285 sires.

Computational times for a sire evaluation that includes dam effects in the model can be projected by combining data in Tables 5 and 6. Table 7 shows projections for 500,000, 800,000, and 1,100,000 records, respectively. The times do not include within herd solution determination for dam EPDs, young animal EPDs, and the time required to compute within herd phenotypic, genetic, and environmental trends.

Considerably fewer records will be deleted from an evaluation when including dam effects in the model. Dams provide within herd indirect ties between sires used in different contemporary groups. Dams are used across contemporary groups within a herd. Therefore, even if a contemporary group has only one sire, the sire will be tied to sires in other contemporary groups since dams are cross classified with

Table 7. Projected computational time requirements

Program	Computation time (CPU min.)[a] Number of weaning weight records 500,000	800,000	1,100,000
Data editing	11	18	25
Dam absorption	7	11	15
Contemporary group absorption	11	17	24
Adding relationships	3	5	7
Equation iteration	67	107	147
Miscellaneous	15	24	33
Total	114	182	251

[a]Assumes a constant number of records for the complete evaluation process and 30,000 sire equations.

contemporary groups. One exception would be the case in which a sire was mated to dams with no other progeny records in the herd (for example, first calf heifers).

Using more performance records and evaluating more of a breed's sires are desirable. However, major computational difficulties could exist if the number of sires increases much beyond 28,000 sires. As the number of sires increases, the time to iterate the evaluation equations increases exponentially.

One solution to the problem of computational difficulty as number of sires increases might be to limit the number of base sires added to the evaluation that have no progeny records. Another possibility would be to delete the oldest birth year group as another birth year group of sires is added.

Computer core requirements are dictated by the new contemporary group absorption program in which three matrices are full-stored in double precision. If the number of contemporary groups per herd does not exceed 160 and the number of sires in that same herd does not exceed 140, then the core requirements will be on the order of 1024K.

One herd in the validation data base had 153 contemporary groups. Each year this number will be increased by at least two and in some cases by three or four. The maximum number of sires used by any one herd was 129. Therefore, it is likely that core requirements will increase beyond 1024K quickly unless some of the early records are deleted when new records are added.

Core requirements are essentially independent of the number of records or the number of herds being considered. External storage requirements for data sets, however, are dependent upon the number of records and herds being analyzed. Most data sets will have to be stored on magnetic tapes due to the sheer volume of data, particularly when within herd estimates and trends are being made. Data storage on tape could slow the evaluations considerably since tape access is slower than disk access.

Sire EPDs for Direct WWT Comparisons

One part of the methodology and computational validation process is the comparison of the proposed evaluation sire EPDs for direct WWT with those obtained in the Angus Sire Evaluation for 1983 (Berger et al., 1983). The comparison requires estimates to be made for the proposed model residual error variance (σ_e^2), dam composite effect variance (σ_d^2),

and sire direct effect variance (σ_s^2). The assumed estimates are 1570 for σ_e^2, 7 for σ_e^2/σ_d^2, and 15 for σ_e^2/σ_s^2.

Three categories of sires are presented in Table 8 to provide a comparison between the two sire evaluation models using Pearson's average product moment correlation and Spearman's rank correlation (SAS, 1982). The correlations are computed for three different categories of effective progeny number (EPN) per sire. The EPN shown here are those obtained with the proposed evaluation. The correlations were also computed for all categories combined together. The categories correspond to high accuracy (>=75 EPN), moderate to high accuracy (>=20 and <75 EPN), and low accuracy (<20 EPN).

In general, the results are what would be expected: the higher the accuracy, the higher the correlation. However, there is less than perfect correlation and there are several possible reasons that include:

1) Average product moment correlations would be less than perfect because the number of performance records on which each sire is evaluated is less with the validation data base.

2) Error to dam and error to sire variance ratios used "best guess" estimates for the validation evaluation. EPDs did tend to be further regressed to the mean for the sires in this evaluation.

3) The validation evaluation may be removing effects of assortative mating that cannot be accounted for in the current sire evaluation. This would affect rank correlations of some sires.

Table 9 shows a list of sires and their EPDs from the 1983 Angus Sire Evaluation for direct WWT and their EPDs from the validation

Table 8. Average product moment and rank-order correlations between 1983 Angus Sire Evaluation and proposed sire evaluation for EPDs for direct WWT

Effective progeny number/sire	Product moment correlation[a]	Rank-order correlation[b]
Sires with >=75	.93	.91
>=20 and <75	.88	.89
<20	.73	.72
All sires	.79	.79

[a]Pearson average product moment correlation.

[b]Spearman rank-order correlation.

evaluation. The top list of sires are evaluated with a very high accuracy in both evaluations. All the sires in this list have an EPN of greater than 86 with the exception of Lundell of Wye, which has 37.5. The rankings compare very well with only one exception. PS Power Play has a 1983 Angus EPD of +40.7 lbs. for direct WWT and only +21.9 lbs. in the validation evaluation. His EPN in the validation evaluation is 180.2.

Until appropriate variance components can be determined for the breed populations, little more can be said about comparisons with the current sire evaluation procedure. The last section of this chapter presents preliminary results associated with the assessment of determining sire prediction accuracy.

Table 9. Actual comparison of selected sires evaluated in 1983 Angus Sire Evaluation and proposed evaluation model

Name of bull	1983 WWT EPD (lbs.)	Rank	Validation EPD (lbs.)	Rank
Sires with high accuracy in the validation evaluation:				
Rito 8221 "Big Moose"	43.3	1	36.4	2
Shoshone Viking GD60	41.3	2	33.9	3
Thomas Chaps	40.8	3	32.5	4
PS Power Play	40.7	4	21.9	11
Shoshone Shannon HC3	40.5	5	36.7	1
Schearbrook Shoshone	37.5	6	26.7	6
Jetliner 707 of Conanga	37.1	7	23.9	9
Black Witch Corbinaire	35.4	8	26.3	7
Rito 149 of Ideal 632 72	31.7	9	29.2	5
Rito 206 of Ideal 2218	31.4	10	22.2	10
Rito 707 of Ideal 836 71	30.2	11	26.3	8
CSU Rito 4114	27.4	12	9.1	17
Band 234 of Ideal 3163	24.2	13	14.7	12
Ankonian Dynamo	23.6	14	10.4	16
Lundell of Wye	22.8	15	14.7	13
Marshall Pride 4956 SAR	22.3	16	12.7	15
CRR Emulous Pacesetter	20.6	17	13.5	14
Black Bull Matt Dillon	8.0	18	- 2.0	18
Emulous 1701 of SAR	5.5	19	- 5.5	19
Sires with low accuracy in validation evaluation:				
Shoshone Shanigan OVK7	53.9		24.9	
Premier Progressor	48.5		26.8	
Mr. Angus	45.6		5.4	
Sayer Patriot	34.0		27.2	

Prediction Accuracy

The purpose of this section is to review a preliminary assessment made on approximations to sire evaluation prediction error variance (PEV). A subset of the validation herds is used to form mixed model sire evaluation equations of order small enough to obtain a direct inverse of the coefficient matrix. Thirteen herds are used and result in a

matrix of size 611 by 611. The matrix includes 582 sires and 18 sire birth year groups. The inversion time was 11 CPU-minutes on Iowa State University's AS-6 computer.

Diagonal elements of the inverse for 13 different sires are given in Table 10 with four approximations to prediction error variance (PEV) divided by σ_e^2. The approximations are similar and differ only in the amount of information included in the denominator term. The approximations are given as follows:

$$\text{Approximation (1)} = 1/\text{EPN}$$

$$\text{Approximation (2)} = 1/(\text{EPN} + \alpha)$$

$$\text{Approximation (3)} = 1/(\text{EPN} + \alpha \cdot a_{ii})$$

$$\text{Approximation (4)} = 1/(\text{EPN} + \alpha \cdot a_{ii(adj)})$$

where

EPN = the i-th sire's effective progeny number,

α = the error variance to sire variance ratio,

a_{ii} = the diagonal element of A^{-1} for the i-th sire.

Effective progeny numbers for the sires range from 1.7 to 256.1. The four approximations are represented in the four columns labeled (1), (2), (3), and (4). Differences for each approximation from the true value are presented in the parentheses.

The magnitude of the differences suggests that one approximation may not be best over the complete range of EPN. Approximations (2) and (4) typically give the best estimates in the low EPN range. Approximation (1) gives the best estimate in the high EPN range.

Table 10. Comparisons of actual prediction error variance (PEV)/σ_e^2 and four approximations

Sire	EPN	PEV/σ_e^2	Approximation (|Diff| from actual) (1) (EPN)	(2) (EPN+α)	(3) (EPN+$\alpha \cdot a_{ii}$)	(4) (EPN+$\alpha a_{ii(adj)}$)
465	1.7	.0955	.5840 (.4885)	.1122 (.0167)	.0821 (.0134)	.0905 (.0050)*[a]
456	5.4	.0818	.1837 (.1019)	.0791 (.0027)*	.0665 (.0153)	.0781 (.0036)
77	9.2	.0327	.1090 (.0763)	.0618 (.0291)	.0174 (.0153)	.0227 (.0100)*
541	10.1	.0556	.0993 (.0437)	.0579 (.0023)*	.0487 (.0069)	.0510 (.0046)
418	12.5	.0491	.0798 (.0307)	.0512 (.0021)*	.0435 (.0056)	.0438 (.0053)
355	23.8	.0346	.0418 (.0072)	.0324 (.0022)*	.0271 (.0075)	.0287 (.0059)
70	35.7	.0517	.0280 (.0237)*	.0234 (.0283)	.0229 (.0288)	.0229 (.0288)
509	39.5	.0271	.0253 (.0018)	.0215 (.0056)	.0200 (.0017)*	.0206 (.0066)
384	39.7	.0279	.0252 (.0027)*	.0214 (.0065)	.0213 (.0066)	.0213 (.0066)
571	40.2	.0263	.0249 (.0014)*	.0212 (.0051)	.0198 (.0065)	.0198 (.0065)
347	57.9	.0201	.0173 (.0028)*	.0154 (.0047)	.0146 (.0055)	.0146 (.0055)
444	133.0	.0120	.0075 (.0045)*	.0071 (.0049)	.0067 (.0053)	.0067 (.0053)
486	256.1	.0092	.0039 (.0053)*	.0038 (.0054)	.0037 (.0055)	.0037 (.0055)
Std. dev. of differences			.1324	.0095	.0071	.0065

[a]Asterisk indicates approximate with smallest difference.

However, the relative magnitudes for all approximations in the high EPN range are similar. The variation in magnitude of the difference from sire to sire is different for the four approximation methods.

The most complex approximation, (4), has the least variation (.0065). This is understandable since approximation (4) includes more relevant information than any of the other approximations. This approximation accounts for the number of progeny records a sire has, the distribution across contemporary groups, an adjusted effect from relative contributions, and the heritability of the trait being measured.

Although not recommended until additional analysis is completed, approximation (4) + .0050 might be a good approximation over the complete range of EPN. In addition, an analysis similar to that of Ufford et al. (1979) needs to be run to determine if a simple regression on EPN gives an appropriate approximation.

The last chapter provides a summary of the major elements of this thesis. The chapter also points out the additional research that must be accomplished before the unification of national beef sire evaluation and within herd evaluations can become a reality.

SUMMARY

The thesis presents a methodology associated with one approach to the unification of national beef sire evaluation and within herd evaluations. A major reason for unification is to provide the basis from which young unsampled yearling bulls can be fairly compared across herds. The unification of the two evaluations provides the means of obtaining best linear unbiased predictors for these young bulls.

Mixed model evaluation methodology is developed for sire direct and maternal weaning weight (WWT) evaluations, dam composite (and separate) direct and maternal WWT evaluations, and young animal evaluations for direct and maternal WWT. Four computer algorithms written in PL/I are developed to support the unification evaluation procedures. Two algorithms are used to absorb within herd dam random effects and contemporary group fixed effects into sire equations for use in the national sire evaluation. The other two algorithms are used for the back solution of within herd composite dam effects and contemporary group estimates after sire expected progeny differences (EPDs) are obtained in the national sire evaluation.

Procedures to determine within herd phenotypic, environmental, and genetic trend lines are presented. Herd trend lines are computed from sire and dam EPDs weighted by their respective numbers of progeny in each calf crop year. Similarly, environmental trend lines are obtained from weighted herd contemporary group estimates. Trend lines provide a breeder with an overall historical assessment of his/her

breeding program as compared with the national average of breeders using artificial insemination.

The methodology and computer algorithms are validated with an actual set of field records provided by the American Angus Association. The records include performance records from 20 purebred Angus herds that, for the most part, have long histories of performance testing.

Sire EPDs obtained from the validation evaluation are compared to sire EPDs obtained from the 1983 Angus Sire Evaluation. The Pearson average product moment correlation between individual sire EPDs over all sires is 0.79. The product moment correlation for high accuracy sires is 0.93. The Spearman rank-order correlation for all sires in the two evaluations is 0.79. The rank-order correlation is 0.91 when considering only the high accuracy sires.

Three reasons may explain why the product moment correlation is not higher than 0.79 when considering all sires. First, both evaluations are done with a floating base. Thus, EPDs are not directly comparable from one evaluation to the other. Secondly, sires in the validation evaluation are evaluated, for the most part, on considerably fewer progeny records. Thirdly, variance ratios used in the validation evaluation are "best guess" estimates, and the EPDs of the validation evaluation tend to be more regressed to the mean than in the 1983 Angus evaluation. The less than perfect rank-order correlation is probably due to the lower number of records available by which to evaluate the sires in the validation evaluation. The rank-order

correlation (0.79) may also be due to the removal of preferential mating effects that are not accounted for in the 1983 Angus evaluation.

The computer algorithms developed for this thesis can be used with minor modification to support operational sire single trait evaluations for birth, weaning, and yearly weights. Using evaluation times obtained from the validation evaluation, the projected computer time to conduct a major evaluation that includes 800,000 performance records is 182 CPU minutes. By comparison, the 1983 Angus Sire Evaluation for weaning weight that had 354,746 records and 28,285 sires took 72 CPU minutes. The majority of the time increase for the proposed evaluation model relates to the contemporary group absorption process. For the absorption, a direct inverse must be computed for contemporary group equations on a herd by herd basis.

Variance component estimation for the traits to be used in mixed model evaluation equations is an area of future research needed to complete the unification of national sire evaluation and within herd evaluations. In addition, two major computer algorithms must be written for the evaluation model. An algorithm is needed for a sire evaluation model for maternal WWT. Another algorithm is needed for the dam evaluation model for direct and maternal WWT effects.

LITERATURE CITED

Berger, P. J. 1984. Current sire evaluation. In Proc. of Workshop on the Prediction of Genetic Values for the Beef Industry, Winrock International, Morrilton, Arkansas.

Berger, P. J., D. E. Wilson, and R. L. Willham. 1983. Angus sire evaluation for 1983. In 1983 Angus Sire Evaluation Report. Amer. Angus Assoc., St. Joseph, MO.

Dempfle, L., C. Hagger, and M. Schneeberger. 1983. On the estimation of genetic parameters via variance components. Genet. Sel. Evol. 15: 425.

Falconer, D. S. 1960. Introduction to quantitative genetics. (Reprint). New York: The Ronald Press Company.

Freeman, A. E. 1980. Impact of sire evaluation on dairy cattle breeding. In Proc. Beef Improvement Federation Res. Symp., Denver, CO.

Guidelines for Uniform Beef Improvement Programs. Program Aid 1020-USDA, Extension Service, 1981. 76 p. Available from United States Department of Agriculture, Washington, D.C.

Hartley, H. O., and J. N. K. Rao. 1967. Maximum likelihood estimation for the mixed analysis of variance model. Biometrika 54:93.

Harville, D. A. 1977. Maximum likelihood approaches to variable component estimation and to related problems. J. Am. Stat. Assoc. 72:320.

Harville, D. A., and C. R. Henderson. 1967. Environmental and genetic trends in production and their effects on sire evaluation. J. Dairy Sci. 50:870.

Henderson, C. R. 1953. Estimation of variance and covariance components. Biometrics 9:226.

Henderson, C. R. 1963. Selection index and expected genetic advance. NAS-NRC 982.

Henderson, C. R. 1966. A sire evaluation method which accounts for unknown genetic and environmental trends, herd differences, season, age effects, and differential culling. In Proc. of Symp. on Estimating Breeding Values of Dairy Sires and Cows, Washington, D.C.

Henderson, C. R. 1973. Sire evaluation and genetic trends. In Proc. Anim. Breeding and Genetics Symp. in honor of Dr. Jay L. Lush. A.S.A.S. and A.S.D.A., Champaign, IL.

Henderson, C. R. 1974. General flexibility of linear model techniques for sire evaluation. J. Dairy Sci. 57:963.

Henderson, C. R. 1975a. Use of relationships among sires to increase accuracy of sire evaluations. J. Dairy Sci. 58:1731.

Henderson, C. R. 1975b. Comparison of alternative sire evaluation methods. J. Anim. Sci. 41:760.

Henderson, C. R. 1975c. Inverse of a matrix of relationships due to sires and maternal grandsires. J. Dairy Sci. 58:1917.

Henderson, C. R. 1975d. Rapid method for computing the inverse of a relationship matrix. J. Dairy Sci. 58:1727.

Henderson, C. R. 1975e. Use of all relatives in intraherd prediction of breeding values and producing abilities. J. Dairy Sci. 58:1910.

Henderson, C. R. 1975f. Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423.

Henderson, C. R. 1980. A simple method for unbiased estimation of variance components in the mixed model. Mimeo. Dept. of Animal Science, Cornell Univ., Ithaca, N.Y.

Henderson, C. R., O. Kempthorne, S. R. Searle, and C. M. von Krosigk. 1959. The estimation of environmental and genetic trends from records subject to culling. Biometrics 15:192.

Hintz, R. L., R. W. Everett, and L. D. Van Vleck. 1978. Estimation of genetic trends from cow and sire evaluations. J. Dairy Sci. 61: 607.

Hudson, G. F. S., and L. D. Van Vleck. 1982. Estimation of components of variance by method 3 and Henderson's new method. J. Dairy Sci. 65:435.

Lush, J. L. 1948. The genetics of populations. Unpublished notes. Dept. of Animal Science, Iowa State Univ., Ames, IA.

Middleton, B. K. 1983. Intra herd prediction of beef cattle genetic values. Ph.D. dissertation. Iowa State Univ., Ames, IA.

Nielson, M. K. 1974. An examination of analysis procedures for national beef sire evaluation programs. Ph.D. dissertation. Iowa State Univ., Ames, IA.

Patterson, H. D., and R. Thompson. 1971. Recovery of interblock information when block sizes are unequal. Biometrika 58:545.

Pollack, E. J., G. R. Ufford, and S. J. Gross. 1977. Comparisons of alternative models for within-herd genetic evaluation of beef cattle. J. Anim. Sci. 45:1010.

Quaas, R. L., and E. J. Pollak. 1980. Mixed model methodology for farm and ranch beef cattle testing programs. J. Anim. Sci. 51:1277.

Rao, C. R. 1971. Minimum variance quadratic unbiased estimation of variance components. J. Multivar. Anal. 1:445.

SAS User's Guide. 1982 edition. Raleigh, North Carolina: SAS Institute, Inc.

Schaeffer, L. R. 1983. Pp. 260-265. In Notes on Linear Model Theory, Best Linear Unbiased Prediction, and Variance Component Estimation. Dept. of Animal and Poultry Science, University of Guelph, Guelph, Ontario.

Schaeffer, L. R., M. G. Freeman, and E. B. Burnside. 1975. Evaluation of Ontario Holstein dairy sires for milk and fat production. J. Dairy Sci. 58:109.

Schaeffer, L. R., and J. W. Wilton. 1975. Predicting beef sire genetic values for growth traits from progeny records. J. Anim. Sci. 41:1592.

Searle, S. R. 1979. Notes on variance component estimation: a detailed account of maximum likelihood and kindred methodology. Paper BU-673-M, Biometrics Unit, Cornell Univ., Ithaca, N.Y.

Slanger, W. D. 1977. An indirect method of forming expressions for the inverses of genetic variance-covariance matrices for two traits. J. Dairy Sci. 60:1440.

Slanger, W. D. 1979. Genetic evaluation of beef cattle for weaning weight. J. Anim. Sci. 48:1070.

Slanger, W. D. 1980. Comparison of within-herd genetic evaluation procedures for beef cattle weaning weight. J. Anim. Sci. 51:595.

Smith, C. 1962. Estimation of genetic change in farm livestock using field records. Animal Prod. 4:239.

Ufford, G. R., C. R. Henderson, and L. D. Van Vleck. 1979. An approximate procedure for determining prediction error variances of sire evaluations. J. Anim. Sci. 62:621.

Van Vleck, L. D., and C. R. Henderson. 1961. Measurement of genetic trend. J. Dairy Sci. 44:1704.

Willham, R. L. 1963. The covariance between relatives for characters composed of components contributed by related individuals. Biometrics 19:18.

Willham, R. L. 1974. Guidelines for national sire evaluation. Guidelines for Uniform Beef Improvement Programs. USDA Extension Service Aid 1020.

Willham, R. L. 1979. Evaluation and direction of beef sire evaluation programs. J. Anim. Sci. 49:592.

Willham, R. L. 1983. What is next in beef sire evaluation? Mimeo. Dept. of Animal Science, Iowa State Univ., Ames, IA.

Willham, R. L., and E. A. Leighton. 1978. Breeding value considerations. In Proc. Beef Improvement Federation Res. Symp., Blacksburg, VA.

Wright, S. 1922. Coefficients of inbreeding and relationships. Amer. Nat. 56:330.

ACKNOWLEDGMENTS

The personal interest and guidance of Dr. R. L. Willham shown to me during the course of my graduate studies in animal breeding is deeply appreciated. I feel that my graduate tenure of five years under his leadership was indeed an honor and I will treasure the memories of the nurturing process that went on.

A special thanks to Dr. P. J. Berger for his counsel and advice given during the thought process and final preparation of this manuscript. My thanks to Drs. Freeman, Spike, and Harville for serving on my doctoral committee. Also, thanks to the people at the American Angus Association and members of the breed improvement committee for their support in pursuing the development of improved beef cattle evaluation procedures. The research reported in this thesis was supported in part by a substantial grant from the American Angus Association.

There is no doubt in my mind that without the support of my family, my desire to return to school and pursue a new career would never have been possible. My wife, Linda, and daughters, Leslie and Katherine, each made many sacrifices over the past five years for which I am deeply grateful. I hope in the future that we can always look back and say it was worth it. In faith we step out into the unknown, but with the assurance that whatever we do will flourish if the final tribute is given to our Lord.

APPENDIX A.

PL/I COMPUTER ALGORITHM FOR INTRAHERD DAM EFFECT ABSORPTION INTO CONTEMPORARY GROUP AND SIRE EFFECTS

```
1.      //ANGUS JOB U3274,WILSON
2.      /*JOBPARM PURGE=NO
7.      //S2 EXEC PLIXCLG,REGION.GO=120K,TIME.GO=5,PARM.GO='ISA(5K)'
8.      //PLI.SYSIN  DD *
9.       ABS_DAM:  PROC OPTIONS(MAIN) REORDER;
10.
11.      /*****************************************************************/
12.      /* Program to absorb dams into contemporary group effects and    */
13.      /* sire effects.  Lines to be changed for different traits       */
14.      /* include:  29, 96/97                                           */
15.      /*****************************************************************/
16.
17.      DCL SYSPRINT FILE STREAM OUTPUT PRINT EXTERNAL;
18.      DCL RECIN INPUT FILE RECORD SEQL BUF ENV(TOTAL
19.          CONSECUTIVE FB RECSIZE(62) BLKSIZE(6200));
20.      DCL RECOUT OUTPUT FILE RECORD SEQL BUF ENV(TOTAL
21.          CONSECUTIVE FB RECSIZE(24) BLKSIZE(6216));
22.
23.      DCL (J,K,CG,TEMP_HERD,TEMP_DAM,TEMP_SIRE,NO_REC,NO_HERD,
24.          NO_HERDREC,NO_DAMS) FIXED BIN(31,0) INIT(0);
25.      DCL (COEF,DAM_REC) FLOAT DEC(16);
26.      DCL (I,CT) FIXED BIN;
27.      DCL EOFSW BIT(1) INIT('1'B);
28.      DCL CONTEMP(20) FIXED BIN(31,0);
29.      DCL ALPHA FLOAT DEC(16) INIT(7);
30.      DCL (TEMP_ADJ,MEAN_REC,SS_TOTAL,SS_CORRECTED) FLOAT DEC(16) INIT(0);
31.      DCL (P,Q,R) PTR;
32.      DCL SIRE_STORE(20,3) CTL FIXED BIN(31,0) INIT((60)0);
33.
34.      DCL 1 REC_IN BASED (P),
35.          2 PAD CHAR(2),
36.          2 HERD CHAR(8),
37.          2 SEX PIC'9',
38.          2 WEAN_DATE PIC'(5)9',
39.          2 WEAN_MGT PIC'9',
40.          2 YEAR_DATE PIC'(5)9',
41.          2 YEAR_TYPE PIC'9',
42.          2 YEAR_MGT PIC'9',
43.          2 DATA_SOURCE PIC'9',
44.          2 SIRE_REG PIC'(8)9',
45.          2 SIRE_YRGP PIC'(2)9',
46.          2 PROGENY_REC PIC'(4)9',
47.          2 DAM_REG PIC'(8)9',
48.          2 DAM_DATE PIC'(5)9',
49.          2 CALF_REG PIC'(8)9',
50.          2 DAM_NO PIC'(2)9';
51.
52.      DCL 1 EQN_OUT BASED (Q),
53.         (2 HERD_OUT,
54.          2 ROW,
55.          2 COL,
56.          2 CODE) FIXED BIN(31,0),
57.          2 COEF_OUT FLOAT DEC(16);
58.
59.      PUT EDIT('*** DAM ABSORPTION - HERD COUNTS ***') (X(48),A(36)) SKIP(5);
60.
61.      ON ENDFILE (RECIN) EOFSW='0'B;
62.
63.      READ FILE (RECIN) SET (P);
64.
```

```
 65.       MEAN_REC=REC_IN.PROGENY_REC;
 66.       NO_REC,NO_HERD = 1;
 67.       SS_TOTAL=REC_IN.PROGENY_REC*REC_IN.PROGENY_REC;
 68.       SS_CORRECTED=0;
 69.
 70.        TEMP_HERD=REC_IN.HERD;
 71.
 72.      /*********************************************************************/
 73.      /* The following do-loop will run through all of the dams within  */
 74.      /* the i-th herd.                                                    */
 75.      /*********************************************************************/
 76.
 77.         DO WHILE (EOFSW & REC_IN.HERD=TEMP_HERD);
 78.
 79.          TEMP_DAM=REC_IN.DAM_REG;
 80.          CT=0;
 81.          ALLOCATE SIRE_STORE(REC_IN.DAM_NO,3);
 82.          COEF=-1/(REC_IN.DAM_NO+ALPHA);
 83.          DAM_REC=0;
 84.          NO_DAMS=NO_DAMS+1;
 85.
 86.      /*********************************************************************/
 87.      /* The following do-loop absorbs the i-th dam into the            */
 88.      /* appropriate contemporary groups within the i-th herd.          */
 89.      /*********************************************************************/
 90.
 91.            DO WHILE (EOFSW & REC_IN.DAM_REG=TEMP_DAM);
 92.
 93.               NO_HERDREC=NO_HERDREC+1;
 94.               CT=CT+1;
 95.               DAM_REC=DAM_REC+REC_IN.PROGENY_REC;
 96.               K=REC_IN.SEX||REC_IN.WEAN_DATE||REC_IN.WEAN_MGT||
 97.                    REC_IN.DATA_SOURCE;
 98.               CONTEMP(CT)=K;
 99.
100.               SIRE_STORE(CT,1)=REC_IN.SIRE_REG;
101.               SIRE_STORE(CT,2)=K;
102.
103.      /*********************************************************************/
104.      /* Determining X'SY, the RHS of the mixed-model equations for     */
105.      /* the contemporary group equations (the first part).             */
106.      /*********************************************************************/
107.
108.                LOCATE EQN_OUT FILE(RECOUT) SET(Q);
109.                HERD_OUT=REC_IN.HERD;
110.                ROW=K;
111.                COL=99999991;
112.                CODE=6;
113.                COEF_OUT=REC_IN.PROGENY_REC;
114.
115.      /* RHS of Z'SY (first part)   */
116.
117.              LOCATE EQN_OUT FILE(RECOUT) SET(Q);
118.                HERD_OUT=TEMP_HERD;
119.                ROW=REC_IN.SIRE_REG;
120.                COL=99999991;
121.                CODE=7;
122.                COEF_OUT=REC_IN.PROGENY_REC;
123.      /*********************************************************************/
124.      /* The following do-loop determines diagonal and off-diagonal     */
```

```
125.      /* elements of the X'SX (cg by cg) matrix as each record for a    */
126.      /* particular dam.  The number of diagonal elements is equal to   */
127.      /* the number of progeny (DAM_NO=n) a dam has.  The number of      */
128.      /* off-diagonal elements is equal to (n**n-n).                    */
129.      /*********************************************************************/
130.
131.                DO I=1 TO CT;
132.
133.                   J=CONTEMP(I);
134.
135.                   IF J¬=K THEN DO;
136.
137.                      LOCATE EQN_OUT FILE(RECOUT) SET(Q);
138.                        HERD_OUT=REC_IN.HERD;
139.                        ROW=J;
140.                        COL=K;
141.                        CODE=2;
142.                        COEF_OUT=COEF;
143.                   END;
144.                   ELSE DO;
145.
146.      L1:             LOCATE EQN_OUT FILE(RECOUT) SET(Q);
147.                        HERD_OUT=REC_IN.HERD;
148.                        ROW,COL=K;
149.                        CODE=1;
150.                        COEF_OUT=1+COEF;
151.
152.                   END;
153.      L2:       END;
154.      L3:       READ FILE (RECIN) SET(P);
155.
156.                IF EOFSW THEN DO;
157.                  NO_REC=NO_REC+1;
158.                  TEMP_ADJ=REC_IN.PROGENY_REC-MEAN_REC;
159.                  SS_TOTAL=SS_TOTAL+REC_IN.PROGENY_REC*
160.                                   REC_IN.PROGENY_REC;
161.                  SS_CORRECTED=SS_CORRECTED+TEMP_ADJ**2-
162.                                        TEMP_ADJ**2/NO_REC;
163.                  MEAN_REC=MEAN_REC+TEMP_ADJ/NO_REC;
164.                END;
165.
166.       END;
167.
168.      /*********************************************************************/
169.      /* The following do-loop determines the elements of X'SZ, the     */
170.      /* contemporary group by sire elements.  This step is             */
171.      /* accomplished after all records for the I-th dam have been      */
172.      /* read.  This is because the number of matings for the k-th      */
173.      /* sire and the I-th dam must be counted.                         */
174.      /*********************************************************************/
175.
176.            DO I=1 TO CT;
177.               TEMP_SIRE=SIRE_STORE(I,1);
178.               SIRE_STORE(I,3)=0;
179.                  DO J=1 TO CT;
180.                     IF SIRE_STORE(J,1)=TEMP_SIRE
181.                              THEN SIRE_STORE(I,3)=SIRE_STORE(I,3)+1;
182.                  END;
183.            END;
184.
```

```
185.      /* Determining the diagonal elements of Z'SZ   */
186.
187.            DO I=1 TO CT;
188.               IF I<CT THEN DO;
189.                  DO J=I+1 TO CT;
190.                    IF SIRE_STORE(J,1)=SIRE_STORE(I,1)
191.                      THEN SIRE_STORE(J,3)=0;
192.                  END;
193.               END;
194.               IF SIRE_STORE(I,3)>0 THEN DO;
195.                  LOCATE EQN_OUT FILE(RECOUT) SET(Q);
196.                    HERD_OUT=TEMP_HERD;
197.                    ROW,COL=SIRE_STORE(I,1);
198.                    CODE=3;
199.                    COEF_OUT=(1+COEF*SIRE_STORE(I,3))*SIRE_STORE(I,3);
200.
201.      /* RHS of Z'SY (second part)   */
202.
203.                  LOCATE EQN_OUT FILE(RECOUT) SET(Q);
204.                    HERD_OUT=TEMP_HERD;
205.                    ROW=SIRE_STORE(I,1);
206.                    COL=99999991;
207.                    CODE=7;
208.                    COEF_OUT=COEF*SIRE_STORE(I,3)*DAM_REC;
209.               END;
210.            END;
211.
212.           DO I=1 TO CT;
213.                    LOCATE EQN_OUT FILE(RECOUT) SET (Q);
214.                      HERD_OUT=TEMP_HERD;
215.                      ROW=CONTEMP(I);
216.                      COL=99999991;
217.                      CODE=6;
218.                      COEF_OUT=COEF*DAM_REC;
219.              DO J=1 TO CT;
220.                 IF I=J THEN DO;
221.      /* X'SZ elements       */
222.                    LOCATE EQN_OUT FILE(RECOUT) SET (Q);
223.                      HERD_OUT=TEMP_HERD;
224.                      ROW=SIRE_STORE(I,2);
225.                      COL=SIRE_STORE(I,1);
226.                      CODE=5;
227.                      COEF_OUT=1+COEF;
228.                 END;
229.                 ELSE DO;
230.                    LOCATE EQN_OUT FILE(RECOUT) SET(Q);
231.                       HERD_OUT=TEMP_HERD;
232.                       ROW=SIRE_STORE(I,2);
233.                       COL=SIRE_STORE(J,1);
234.                       CODE=5;
235.                       COEF_OUT=COEF;
236.
237.      /*  Determining off-diagonal elements of Z'SZ    */
238.
239.                 IF SIRE_STORE(I,1)=SIRE_STORE(J,1) THEN
240.                      GO TO L6;
241.                    LOCATE EQN_OUT FILE(RECOUT) SET(Q);
242.                       HERD_OUT=TEMP_HERD;
243.                       ROW=SIRE_STORE(I,1);
244.                       COL=SIRE_STORE(J,1);
```

```
245.                        CODE=4;
246.                        COEF_OUT=COEF;
247.      L6:          END;
248.                 END;
249.           END;
250.
251.           FREE SIRE_STORE;
252.           IF TEMP_HERD¬=REC_IN.HERD THEN DO;
253.             NO_HERD=NO_HERD+1;
254.             PUT LIST('HERD ID:',TEMP_HERD,'NO. RECORDS:',NO_HERDREC,
255.                 'NO. DAMS:',NO_DAMS) SKIP;
256.             TEMP_HERD=REC_IN.HERD;
257.             NO_HERDREC=0;
258.             NO_DAMS=0;
259.           END;
260.           END;
261.
262.             PUT LIST('HERD ID:',TEMP_HERD,'NO. RECORDS:',NO_HERDREC,
263.                 'NO. DAMS:',NO_DAMS) SKIP;
264.             PUT EDIT('SUMMARY COUNTS AND STATISTICS')
265.                (X(40),A(29)) SKIP(2);
266.             PUT EDIT('NUMBER OF HERDS =        ',NO_HERD)
267.                 (A(24),X(8),F(5)) SKIP;
268.             PUT EDIT('NUMBER OF RECORDS =      ',NO_REC)
269.                 (A(24),X(6),F(7)) SKIP;
270.             PUT EDIT('PHENOTYPIC AVERAGE =     ',MEAN_REC)
271.                 (A(24),X(6),F(7,2)) SKIP;
272.             PUT EDIT('TOTAL SS (y"y) =         ',SS_TOTAL)
273.                 (A(24),F(14,2)) SKIP;
274.             PUT EDIT('SS - CORRECTED =         ',SS_CORRECTED)
275.                 (A(24),F(14,2)) SKIP;
276.             PUT EDIT('PHENOTYPIC VARIANCE =    ',SS_CORRECTED/(NO_REC-1))
277.                 (A(24),F(13,2)) SKIP;
278.
279.       END ABS_DAM;
280.      /*
281.      //GO.RECIN DD DSN=A.U3274.WREC2,UNIT=3330,DISP=(OLD,KEEP),
282.      // VOL=SER=AGS208
283.      //GO.RECOUT DD DSN=&&TEMP1,UNIT=SCRTCH,DISP=(NEW,PASS),
284.      // SPACE=(TRK,(50,50),RLSE),
285.      // DCB=(RECFM=FB,LRECL=24,BLKSIZE=6216)
286.      //ST3 EXEC MOD,PACK=AGS208
287.      //MOD.SYSIN DD *
288.        SCRATCH DSNAME=A.U3274.ABSSORT,VOL=DISK=AGS208
289.      /*
290.      //STEP4 EXEC SYMSORT,TIME=5,TRACKS=900
291.      //SYSIN  DD *
292.        SORT FIELDS=(1,4,BI,A,13,4,BI,A,5,4,BI,A,9,4,BI,A)
293.        SUM FIELDS=(17,8,FL)
294.      //SORTIN DD DSN=&&TEMP1,UNIT=SCRTCH,DISP=(OLD,DELETE)
296.      //SORTOUT DD DSN=A.U3274.ABSSORT,UNIT=3330,VOL=SER=AGS208,
297.      // DISP=(NEW,CATLG),SPACE=(TRK,(50,50),RLSE),
298.      // DCB=(RECFM=FB,LRECL=24,BLKSIZE=6216)
```

APPENDIX B.

PL/I COMPUTER ALGORITHM FOR CONTEMPORARY GROUP EFFECT ABSORPTION INTO SIRE EQUATIONS

```
1.    //ANGUS JOB U3274,WILSON
2.    //S1 EXEC MOD,PACK=AGS208
3.    //MOD.SYSIN DD *
4.      SCRATCH DSNAME=A.U3274.CONTBCK,VOL=DISK=AGS208
7.      SCRATCH DSNAME=A.U3274.FULLWEQN,VOL=DISK=AGS208
8.      SCRATCH DSNAME=A.U3274.BULLIST,VOL=DISK=AGS208
9.    /*
10.   //S2 EXEC PLIXCLG,REGION.GO=1024K,TIME.GO=15,PARM.GO='ISA(820K)'
11.   //PLI.SYSIN  DD *
12.    ABS_GP:  PROC OPTIONS(MAIN) REORDER;
13.
14.    /*****************************************************************/
15.    /* Program to determine the inverse of X'SX and then absorb      */
16.    /* contemporary group effects into sire effects.                 */
17.    /* There are five basic data sets generated in this program:     */
18.    /* (1)  full-stored Z'SZ and Z'Sy equations, (2)  the inverse     */
19.    /* of X'SX for back solution of herd contemporary group effects, */
20.    /* (3) elements of X'Sy for back solution, (4) elements of X'SX  */
21.    /* for back solution, and (5) a list of bulls used across the    */
22.    /* herds.                                                        */
23.    /*****************************************************************/
24.
25.    DCL SYSPRINT FILE STREAM OUTPUT PRINT EXTERNAL;
26.    DCL RECIN INPUT FILE RECORD SEQL BUF ENV(TOTAL
27.        CONSECUTIVE FB RECSIZE(24) BLKSIZE(6216));
28.    DCL (RECOUT,BULLIST) OUTPUT FILE RECORD SEQL BUF ENV(TOTAL
29.        CONSECUTIVE FB RECSIZE(20) BLKSIZE(6220));
30.    DCL SAVEOUT OUTPUT FILE RECORD SEQL BUF ENV(TOTAL
31.        CONSECUTIVE FB RECSIZE(24) BLKSIZE(6216));
32.
33.    DCL SUBSTR BUILTIN;
34.    DCL (I,J,K,L,N,#CGS,#SIRES,NO_HERDS) FIXED BIN INIT(0);
35.    DCL (TEMP_HERD,SIRE_VEC(150) INIT((150)0),
36.      CG_VEC(300) INIT((300)0))  FIXED BIN(31,0);
37.    DCL EOFSW BIT(1) INIT('1'B);
38.    DCL (TEMP,TEMPVEC(#CGS) CTL,BETA,SUM) FLOAT DEC(16);
39.    DCL SOURCE CHAR(1);
40.    DCL (XSX(#CGS,#CGS) CTL,ZSZ(#SIRES,#SIRES)
41.        CTL,CG_VALUE(300) INIT((300)0),
42.        SIRE_VALUE(150) INIT((150)0)) FLOAT DEC(16);
43.    DCL (XSZ(#CGS,#SIRES) CTL, ZSX(#SIRES,#CGS) CTL,
44.        T(#SIRES,#CGS) CTL) FLOAT DEC(16);
45.    DCL (XSY(#CGS) CTL,ZSY(#SIRES) CTL) FLOAT DEC(16);
46.
47.    DCL (P,Q,R) PTR;
48.
49.    DCL 1 REC_IN BASED (Q),
50.       (2 HERD,
51.        2 ROW,
52.        2 COL,
53.        2 CODE) FIXED BIN(31,0),
54.        2 COEF FLOAT DEC(16);
55.
56.    DCL 1 REC_OUT BASED (R),
57.       (2 HERD_OUT,
58.        2 ROW_OUT,
59.        2 COL_OUT) FIXED BIN(31,0),
60.        2 COEF_OUT FLOAT DEC(16);
61.
62.    DCL 1 SAVE_OUT BASED (P),
```

```
63.          (2 HERD_SAVE,
64.           2 ROW_SAVE,
65.           2 COL_SAVE,
66.           2 CODE_SAVE) FIXED BIN(31,0),
67.           2 COEF_SAVE FLOAT DEC(16);
68.
69.       PUT EDIT('***  CONTEMPORARY GROUP ABSORPTION - COUNTS  ***')
70.          (X(38),A(48)) SKIP(5);
71.       PUT SKIP(5);
72.       ON ENDFILE(RECIN) EOFSW='0'B;
73.
74.       READ FILE (RECIN) SET (Q);
75.
76.       DO WHILE (EOFSW);
77.       TEMP_HERD=REC_IN.HERD;
78.       N=0;
79.       K=0;
80.
81.       DO WHILE (EOFSW & REC_IN.HERD=TEMP_HERD);
82.
83.       /*********************************************************************/
84.       /* Reading diagonal elements X'SX and determining the number of   */
85.       /* contemporary groups (#cgs).                                       */
86.       /*********************************************************************/
87.
88.          DO WHILE (REC_IN.CODE = 1);
89.             N=N+1;
90.             SOURCE=SUBSTR(REC_IN.ROW,31,1);
91.             CG_VEC(N)=REC_IN.ROW;
92.             CG_VALUE(N)=REC_IN.COEF;
93.             READ FILE(RECIN) SET (Q);
94.          END;
95.          #CGS=N;
96.
97.       /*********************************************************************/
98.       /* Putting diagonal elements into X'SX matrix.                       */
99.       /*********************************************************************/
100.
101.         ALLOCATE XSX INIT((#CGS*#CGS)0);
102.
103.         DO I=1 TO #CGS;
104.            XSX(I,I)=CG_VALUE(I);
105.         END;
106.
107.      /*********************************************************************/
108.      /* Reading in the rest of X'SX off-diagonal elements.                */
109.      /*********************************************************************/
110.
111.         DO WHILE (REC_IN.CODE = 2);
112.            I=FIND(REC_IN.ROW,1,#CGS,1);
113.            J=FIND(REC_IN.COL,1,#CGS,1);
114.            XSX(I,J),XSX(J,I)=REC_IN.COEF;
115.            READ FILE(RECIN) SET(Q);
116.         END;
117.
118.      /*********************************************************************/
119.      /* Routine to determine the inverse of X'SX.                         */
120.      /*********************************************************************/
121.
122.         L=#CGS-1;
```

```
123.          ALLOCATE TEMPVEC INIT((#CGS)0);
124.
125.          DO I=1 TO #CGS;
126.             TEMP=XSX(1,1);
127.                DO J=1 TO L;
128.                   TEMPVEC(J)=XSX(1,J+1)/TEMP;
129.                END;
130.             TEMPVEC(#CGS)=1.0/TEMP;
131.                DO J=1 TO L;
132.                   TEMP=XSX(J+1,1);
133.                      DO K=1 TO L;
134.                         XSX(J,K)=XSX(J+1,K+1)-(TEMP*TEMPVEC(K));
135.                      END;
136.                   XSX(J,#CGS)=(-TEMP)*TEMPVEC(#CGS);
137.                END;
138.                DO J=1 TO #CGS;
139.                   XSX(#CGS,J)=TEMPVEC(J);
140.                END;
141.          END;
142.
143.       /*************************************************************************/
144.       /* Store elements of inv(X'SX) for back solution of contemporary  */
145.       /* group effects.  Matrix is half-stored (upper).  Code=1 for     */
146.       /* diagonal elements, code=2 for the off-diagonal elements.       */
147.       /*************************************************************************/
148.
149.          DO I=1 TO #CGS;
150.             DO J=I TO #CGS;
151.                LOCATE SAVE_OUT FILE(SAVEOUT) SET (P);
152.                  HERD_SAVE=TEMP_HERD;
153.                  ROW_SAVE=CG_VEC(I);
154.                  COL_SAVE=CG_VEC(J);
155.                  IF I=J THEN CODE_SAVE=1;
156.                     ELSE CODE_SAVE=2;
157.                  COEF_SAVE=XSX(I,J);
158.             END;
159.          END;
160.
161.       /*************************************************************************/
162.       /* Reading diagonal elements Z'SZ and determining the number of   */
163.       /* sires (#sires).                                                */
164.       /*************************************************************************/
165.
166.          N=0;
167.          DO WHILE (REC_IN.CODE = 3);
168.             N=N+1;
169.             SIRE_VEC(N)=REC_IN.ROW;
170.             SIRE_VALUE(N)=REC_IN.COEF;
171.             READ FILE(RECIN) SET (Q);
172.          END;
173.          #SIRES=N;
174.
175.       /*************************************************************************/
176.       /* Putting diagonal elements into Z'SZ matrix.                    */
177.       /*************************************************************************/
178.
179.          ALLOCATE ZSZ INIT((#SIRES*#SIRES)0);
180.
181.          DO I=1 TO #SIRES;
182.             ZSZ(I,I)=SIRE_VALUE(I);
```

```
183.          END;
184.
185.       /*********************************************************************/
186.       /* Reading in the rest of Z'SZ off-diagonal elements.                */
187.       /*********************************************************************/
188.
189.          DO WHILE (REC_IN.CODE = 4);
190.             I=FIND(REC_IN.ROW,1,#SIRES,2);
191.             J=FIND(REC_IN.COL,1,#SIRES,2);
192.             ZSZ(I,J),ZSZ(J,I)=REC_IN.COEF;
193.             READ FILE(RECIN) SET(Q);
194.          END;
195.
196.          ALLOCATE XSZ INIT((#CGS*#SIRES)0),ZSX INIT((#CGS*#SIRES)0);
197.
198.          DO WHILE (REC_IN.CODE = 5);
199.             I=FIND(REC_IN.ROW,1,#CGS,1);
200.             J=FIND(REC_IN.COL,1,#SIRES,2);
201.              XSZ(I,J),ZSX(J,I)=REC_IN.COEF;
202.              READ FILE(RECIN) SET (Q);
203.           END;
204.
205.       /*********************************************************************/
206.       /* Store elements of X'SZ for back solution of cg effects.           */
207.       /*********************************************************************/
208.
209.          DO I=1 TO #CGS;
210.             DO J=1 TO #SIRES;
211.                LOCATE SAVE_OUT FILE(SAVEOUT) SET (P);
212.                  HERD_SAVE=TEMP_HERD;
213.                  ROW_SAVE=CG_VEC(I);
214.                  COL_SAVE=SIRE_VEC(J);
215.                  CODE_SAVE=3;
216.                  COEF_SAVE=XSZ(I,J);
217.             END;
218.          END;
219.          ALLOCATE XSY INIT((#CGS)0);
220.
221.          DO WHILE (REC_IN.CODE = 6);
222.             I=FIND(REC_IN.ROW,1,#CGS,1);
223.              XSY(I)=REC_IN.COEF;
224.             READ FILE(RECIN) SET (Q);
225.          END;
226.
227.       /*********************************************************************/
228.       /* Store elements of X'SY for back solution of cg effects.           */
229.       /*********************************************************************/
230.
231.          DO I=1 TO #CGS;
232.                LOCATE SAVE_OUT FILE(SAVEOUT) SET (P);
233.                  HERD_SAVE=TEMP_HERD;
234.                  ROW_SAVE=CG_VEC(I);
235.                  COL_SAVE=99999991;
236.                  CODE_SAVE=4;
237.                  COEF_SAVE=XSY(I);
238.          END;
239.          ALLOCATE ZSY INIT((#SIRES)0);
240.
241.          DO WHILE (EOFSW & REC_IN.CODE = 7);
242.             I=FIND(REC_IN.ROW,1,#SIRES,2);
```

```
243.            ZSY(I)=REC_IN.COEF;
244.            READ FILE(RECIN) SET (Q);
245.          END;
246.
247.       END;
248.
249.       /*********************************************************************/
250.       /* Computing and intermediate T(I,J)=Z'SX(X'SX+BETA) ¬1.             */
251.       /*********************************************************************/
252.
253.         ALLOCATE T INIT((#SIRES*#CGS)0);
254.
255.         DO I=1 TO #SIRES;
256.            DO J=1 TO #CGS;
257.               DO K=1 TO #CGS;
258.                  T(I,J)=T(I,J)+ZSX(I,K)*XSX(K,J);
259.               END;
260.            END;
261.         END;
262.
263.       /*********************************************************************/
264.       /* Computing the completely absorbed set of sire equations.          */
265.       /* There equations are full stored.                                  */
266.       /*********************************************************************/
267.
268.         DO I=1 TO #SIRES;
269.            DO J=1 TO #SIRES;
270.               SUM=0;
271.               DO K=1 TO #CGS;
272.                  SUM=SUM+T(I,K)*XSZ(K,J);
273.               END;
274.               ZSZ(I,J)=ZSZ(I,J)-SUM;
275.               LOCATE REC_OUT FILE(RECOUT) SET(R);
276.                 HERD_OUT=22;
277.                 ROW_OUT=SIRE_VEC(I);
278.                 COL_OUT=SIRE_VEC(J);
279.                 COEF_OUT=ZSZ(I,J);
280.               IF J=I THEN DO;
281.                 LOCATE REC_OUT FILE(BULLIST) SET(R);
282.                   HERD_OUT=22;
283.                   ROW_OUT=SIRE_VEC(I);
284.                   COL_OUT=1;
285.                   COEF_OUT=ZSZ(I,J);
286.               END;
287.            END;
288.         END;
289.
290.       /*********************************************************************/
291.       /* Computing completely absorbed RHS side of the sire equations.  */
292.       /*********************************************************************/
293.
294.         DO I=1 TO #SIRES;
295.            SUM=0;
296.            DO J=1 TO #CGS;
297.               SUM=SUM+T(I,J)*XSY(J);
298.            END;
299.            ZSY(I)=ZSY(I)-SUM;
300.
301.               LOCATE REC_OUT FILE(RECOUT) SET(R);
302.                 HERD_OUT=22;
```

```
303.                ROW_OUT=SIRE_VEC(I);
304.                COL_OUT=99999991;
305.                COEF_OUT=ZSY(I);
306.
307.          END;
308.
309.          FREE TEMPVEC,XSX,XSY,ZSZ,ZSX,XSZ,ZSY,T;
310.
311.        PUT LIST('HERD ID:  ',TEMP_HERD,'NO. CGS:  ',
312.              #CGS,'NO. SIRES:  ',#SIRES) SKIP;
313.        NO_HERDS=NO_HERDS+1;
314.        END;
315.
316.        PUT LIST ('TOTAL NUMBER OF HERDS:  ',NO_HERDS) SKIP(5);
317.
318.        /*********************************************************************/
319.        /* Routine to find the row of CG_VEC(I) that a particular         */
320.        /* contemporary group is in.                                      */
321.        /*********************************************************************/
322.
323.        FIND:  PROC(ID,START,STOP,SW) RETURNS(FIXED BIN) REORDER;
324.           DCL (START,STOP,SW,HIGH,LOW,MID,MATCH INIT(0)) FIXED BIN;
325.           DCL ID FIXED BIN(31,0);
326.
327.           LOW=START;
328.           HIGH=STOP;
329.
330.           IF SW=1 THEN DO;
331.           DO WHILE (HIGH>=LOW & MATCH=0);
332.             MID=(LOW+HIGH)/2;
333.             SELECT;
334.             WHEN (ID=CG_VEC(MID)) MATCH=MID;
335.             WHEN (ID>CG_VEC(MID)) LOW=MID+1;
336.             WHEN (ID<CG_VEC(MID)) HIGH=MID-1;
337.              END;
338.           END;
339.           END;
340.           ELSE DO;
341.           DO WHILE (HIGH>=LOW & MATCH=0);
342.             MID=(LOW+HIGH)/2;
343.             SELECT;
344.             WHEN (ID=SIRE_VEC(MID)) MATCH=MID;
345.             WHEN (ID>SIRE_VEC(MID)) LOW=MID+1;
346.             WHEN (ID<SIRE_VEC(MID)) HIGH=MID-1;
347.              END;
348.           END;
349.           END;
350.           RETURN (MATCH);
351.         END FIND;
352.         END ABS_GP;
353.       /*
354.       //GO.RECIN DD DSN=A.U3274.ABSSORT,UNIT=3330,DISP=(OLD,KEEP),
355.       // VOL=SER=AGS208
356.       //GO.RECOUT DD DSN=&&TEMP1,UNIT=SCRTCH,DISP=(NEW,PASS),
357.       // SPACE=(TRK,(50,50),RLSE),
358.       // DCB=(RECFM=FB,LRECL=20,BLKSIZE=6220)
359.       //GO.BULLIST DD DSN=&&TEMP2,UNIT=SCRTCH,DISP=(NEW,PASS),
360.       // SPACE=(TRK,(50,50),RLSE),
361.       // DCB=(RECFM=FB,LRECL=20,BLKSIZE=6220)
362.       //GO.SAVEOUT DD DSN=A.U3274.CONTBCK,UNIT=3330,DISP=(NEW,CATLG),
```

```
363.      // SPACE=(TRK,(50,50),RLSE),VOL=SER=AGS208,
364.      // DCB=(RECFM=FB,LRECL=24,BLKSIZE=6216)
365.      //STEP3 EXEC SYMSORT,TIME=3,TRACKS=700
366.      //SYSIN DD *
367.        SORT FIELDS=(5,8,BI,A)
368.        SUM FIELDS=(13,8,FL)
369.      //SORTIN DD DSN=&&TEMP1,UNIT=SCRTCH,DISP=(OLD,DELETE)
371.      //SORTOUT DD DSN=A.U3274.FULLWEQN,UNIT=3330,VOL=SER=AGS208,
372.      // DISP=(NEW,CATLG),SPACE=(TRK,(50,50),RLSE),
373.      // DCB=(RECFM=FB,LRECL=20,BLKSIZE=6220)
374.      //STEP4 EXEC SYMSORT,TIME=2,TRACKS=400
375.      //SYSIN DD *
376.        SORT FIELDS=(5,4,BI,A)
377.        SUM FIELDS=(9,4,BI,13,8,FL)
378.      //SORTIN DD DSN=&&TEMP2,UNIT=SCRTCH,DISP=(OLD,DELETE)
380.      //SORTOUT DD DSN=A.U3274.BULLIST,UNIT=3330,VOL=SER=AGS208,
381.      // DISP=(NEW,CATLG),SPACE=(TRK,(10,10),RLSE),
382.      // DCB=(RECFM=FB,LRECL=20,BLKSIZE=6220)
```

APPENDIX C.

PL/I COMPUTER ALGORITHM FOR BACK SOLUTION OF CONTEMPORARY GROUP ESTIMATES

```
1.      //WILSON JOB U3274,WILSON
1.1     /*JOBPARM LINES=2
2.      //ST1 EXEC MOD,PACK=AGS208
3.      //MOD.SYSIN DD *
4.        SCRATCH DSNAME=A.U3274.CONTSOL,VOL=DISK=AGS208
5.      /*
6.      //ST2 EXEC PLIXCLG,REGION.GO=640K,TIME.GO=5,PARM.GO='ISA(360K)'
7.      //PLI.SYSIN DD *
8.       CGP_SOL: PROC OPTIONS(MAIN) REORDER;
9.       DCL SYSPRINT FILE STREAM OUTPUT PRINT EXTERNAL;
10.      DCL SOLIN INPUT FILE RECORD SEQL BUF
10.5           ENV(CONSECUTIVE VBS RECSIZE(56) BLKSIZE(6233));
12.      DCL SAVEIN INPUT FILE RECORD SEQL
12.5           BUF ENV(CONSECUTIVE FB
13.        RECSIZE(24) BLKSIZE(6216));
13.1     DCL CONTSOL RECORD FILE RECORD SEQL BUF
13.15          ENV(CONSECUTIVE FB RECSIZE(20)
13.65            BLKSIZE(6220));
14.      DCL (#TSIRES,#HERDS,#CGS,#SIRES) FIXED BIN INIT(0);
15.      DCL (SUM INIT(0),INVXSX(160,160)
15.5           INIT((25600)0),XSZ(160,140)
16.        INIT((22400)0), XSY(160) INIT((160)0),XSY_ADJ(160) INIT((160)0))
17.        FLOAT DEC(16);
17.1     DCL SIREHD_LIST(140) INIT((140)0) FIXED BIN(31,0);
17.2     DCL (P,Q,R) PTR;
17.3     DCL (I,J,K,L INIT(0)) FIXED BIN;
17.4     DCL (TEMP_HERD,TEMP_ROW INIT(0)) FIXED BIN(31,0);
17.5     DCL EOFSW BIT(1) INIT('1'B);
17.6     DCL EOFSW2 BIT(1) INIT('1'B);
18.
19.      DCL 1 SOL_IN BASED(Q),
20.         (2 REG,
21.          2 BYR,
22.          2 EQN#) FIXED BIN(31,0),
23.         (2 RSQ1,
24.          2 EPN,
25.          2 SHAT,
26.          2 UHAT,
27.          2 RSQ2) FLOAT DEC(16);
28.
29.      DCL 1 SAVE_OUT BASED(R),
30.         (2 PAD,
31.          2 HERD,
32.          2 CG) FIXED BIN(31,0),
33.          2 CGSOL FLOAT DEC(16);
34.
35.      DCL 1 SIRE_LIST(1000),
36.          2 REG INIT((1000)0) FIXED BIN(31,0),
37.          2 SOLUTION INIT((1000)0) FLOAT DEC(16);
38.
39.      DCL 1 CG_LIST(160),
40.          2 CG INIT((160)0) FIXED BIN(31,0);
41.
42.      DCL 1 SAVE_IN BASED(P),
43.         (2 HERD,
44.          2 ROW,
45.          2 COL,
46.          2 CODE) FIXED BIN(31,0),
47.          2 COEF FLOAT DEC(16);
48.
```

```
49.       /* Read in sire solutions.                                          */
50.
51.       ON ENDFILE(SOLIN) EOFSW='0'B;
52.       READ FILE(SOLIN) SET(Q);
53.
54.       DO WHILE(EOFSW);
55.         #TSIRES=#TSIRES+1;
56.         SIRE_LIST.REG(#TSIRES)=SOL_IN.REG;
57.         SIRE_LIST.SOLUTION(#TSIRES)=SOL_IN.UHAT;
58.         READ FILE(SOLIN) SET(Q);
59.       END;
60.
61.       /* Read in inv(X'SX) elements.                                      */
62.
64.       ON ENDFILE(SAVEIN) EOFSW2='0'B;
65.       READ FILE(SAVEIN) SET(P);
66.
67.       DO WHILE(EOFSW2);
68.         #HERDS=#HERDS+1;
69.         TEMP_HERD=SAVE_IN.HERD;
70.
71.           DO WHILE(EOFSW2 & TEMP_HERD=SAVE_IN.HERD);
72.             J,I=0;
73.               DO WHILE(SAVE_IN.CODE=1 | SAVE_IN.CODE=2);
74.                 IF SAVE_IN.CODE=1 THEN DO;
75.                   I=I+1;
76.                   INVXSX(I,I)=SAVE_IN.COEF;
77.                   CG_LIST.CG(I)=SAVE_IN.ROW;
78.                   J=I;
79.                 END;
80.                 ELSE DO;
81.                   J=J+1;
82.                   INVXSX(I,J),INVXSX(J,I)=SAVE_IN.COEF;
83.                 END;
83.01
83.02               READ FILE(SAVEIN) SET(P);
83.03
83.1              END;
84.
85.                 #CGS=I;
86.
87.       /*  Read in X'SZ elements.                                          */
88.
89.                 TEMP_ROW=SAVE_IN.ROW;
90.                 J,I=0;
91.
92.                 DO WHILE(SAVE_IN.CODE=3);
93.                   I=I+1;
94.                     DO WHILE(TEMP_ROW=SAVE_IN.ROW);
95.                       J=J+1;
96.                       XSZ(I,J)=SAVE_IN.COEF;
97.                       IF I=1 THEN STREHD_LIST(J)=SAVE_IN.COL;
98.                       READ FILE(SAVEIN) SET(P);
99.                     END;
100.
100.1                    IF I=1 THEN #SIRES=J;
101.                    TEMP_ROW=SAVE_IN.ROW;
101.1                   J=0;
102.
103.                END;
```

```
104.
108.      /* Read in X'SY elements.                                          */
109.
109.1               I=0;
110.               DO WHILE(EOFSW2 & SAVE_IN.CODE=4);
111.                 I=I+1;
112.                 XSY(I)=SAVE_IN.COEF;
112.1                READ FILE(SAVEIN) SET(P);
113.               END;
114.
115.      /* Print herd counts for a check on program operation.     */
116.
119.               PUT EDIT('HERD ID = ',TEMP_HERD,'NO. OF CGS = ',#CGS,
119.1                      'NO. OF SIRES = ',#SIRES) (A(15),X(2),F(8),X(2),
119.2                     A(15),X(2),F(8),X(2),A(15),X(2),F(8)) SKIP;
120.
121.      /* Compute contemporary group solutions, where solutions  */
122.      /* are given by:  c¢ = inv(X'SX) *  X'SY-X'SZ*S¢(star)    */
124.      /* Compute X'SY - X'SZ * S¢ first,      call XSY_A DJ(J)       */
125.
126.               DO I=1 TO #CGS;
127.                 SUM=0;
128.                   DO J=1 TO #SIRES;
129.                     K=FIND(SIREHD_LIST(J),1,#TSIRES,1);
130.                     SUM=SUM+XSZ(I,J)*SIRE_LIST.SOLUTION(K);
131.                   END;
132.                 XSY_ADJ(I)=XSY(I)-SUM;
133.               END;
134.
135.      /* Compute and store contemporary group solutions.   */
136.
137.               DO I=1 TO #CGS;
138.                 SUM=0;
139.                   DO J=1 TO #CGS;
140.                     SUM=SUM+INVXSX(I,J)*XSY_ADJ(J);
141.                   END;
142.
143.                   LOCATE SAVE_OUT FILE(CONTSOL) SET(R);
144.                     SAVE_OUT.PAD=22;
145.                     SAVE_OUT.HERD=TEMP_HERD;
146.                     SAVE_OUT.CG=CG_LIST.CG(I);
147.                     SAVE_OUT.CGSOL=SUM;
148.                END;
149.
150.      /* Continue with the next herd or stop if EOFSW='0'B.  */
151.
151.1          END;
153.      END;
154.      PUT LIST('TOTAL NUMBER OF HERDS = ',#HERDS) SKIP;
155.      PUT LIST('TOTAL NUMBER OF SIRES = ',#TSIRES) SKIP;
156.
157.
158.      FIND: PROC(ID,START,STOP,SW) RETURNS(FIXED BIN) REORDER;
159.        DCL (START,STOP,SW,HIGH,LOW,MID,MATCH INIT(0)) FIXED BIN;
160.        DCL ID FIXED BIN(31,0);
161.        LOW=START;
162.        HIGH=STOP;
163.
164.        IF SW=1 THEN DO;
165.           DO WHILE(HIGH>=LOW & MATCH=0);
```

```
165.1             MID=(LOW+HIGH)/2;
166.              SELECT;
167.                WHEN (ID=SIRE_LIST.REG(MID)) MATCH=MID;
168.                WHEN (ID>SIRE_LIST.REG(MID)) LOW=MID+1;
169.                WHEN (ID<SIRE_LIST.REG(MID)) HIGH=MID-1;
170.              END;
171.           END;
172.         END;
173.
174.       RETURN (MATCH);
175.      END FIND;
176.      END CGP_SOL;
177.     /*
178.     //GO.SOLIN DD DSN=A.U3274.WACCLIST,UNIT=3330,DISP=(OLD,KEEP),
179.     // VOL=SER=AGS208
180.     //GO.SAVEIN DD DSN=A.U3274.CONTBCK,UNIT=3330,DISP=(OLD,KEEP),
181.     // VOL=SER=AGS208
182.     //GO.CONTSOL DD DSN=A.U3274.CONTSOL,UNIT=3330,DISP=(NEW,CATLG),
183.     // DCB=(RECFM=FB,LRECL=20,BLKSIZE=6220),VOL=SER=AGS208,
184.     // SPACE=(TRK,(20,20),RLSE)
185.     //
```

APPENDIX D.

PL/I COMPUTER ALGORITHM FOR BACK SOLUTION OF WITHIN HERD DAM PREDICTORS AND ESTIMATION OF MODEL ERROR VARIANCE

```
1.      //WILSON JOB U3274,WILSON
2.      /*JOBPARM LINES=2
3.      //ST1 EXEC MOD,PACK=AGS208
4.      //MOD.SYSIN DD *
5.        SCRATCH DSNAME=A.U3274.DAMSOL,VOL=DISK=AGS208
6.        SCRATCH DSNAME=A.U3274.CGTRD,VOL=DISK=AGS208
7.        SCRATCH DSNAME=A.U3274.DAMTRD,VOL=DISK=AGS208
8.        SCRATCH DSNAME=A.U3274.SIRETRD,VOL=DISK=AGS208
9.        SCRATCH DSNAME=A.U3274.PHTRD,VOL=DISK=AGS208
9.1       SCRATCH DSNAME=A.U3274.PHTRD2,VOL=DISK=AGS208
10.     /*
11.     //ST2 EXEC PLIXCLG,REGION.GO=420K,TIME.GO=5,PARM.GO='ISA(28K)'
12.     //PLI.SYSIN DD *
13.      SS_SQRS: PROC OPTIONS(MAIN) REORDER;
14.      DCL SYSPRINT FILE STREAM OUTPUT PRINT EXTERNAL;
15.      DCL BYRIN INPUT FILE RECORD SEQL BUF ENV(
16.          CONSECUTIVE VBS RECSIZE(32) BLKSIZE(13024));
17.      DCL SIREIN INPUT FILE RECORD SEQL BUF ENV(
18.          CONSECUTIVE VBS RECSIZE(56) BLKSIZE(6233));
19.      DCL CGIN INPUT FILE RECORD SEQL BUF ENV(TOTAL
20.          CONSECUTIVE FB RECSIZE(20) BLKSIZE(6220));
21.      DCL DATAIN INPUT FILE RECORD SEQL BUF ENV(TOTAL
22.          CONSECUTIVE FB RECSIZE(62) BLKSIZE(13020));
23.      DCL RECOUT OUTPUT FILE RECORD SEQL BUF ENV(
24.          CONSECUTIVE FB RECSIZE(28) BLKSIZE(6216));
25.      DCL TREND OUTPUT FILE RECORD SEQL BUF ENV(TOTAL
26.          CONSECUTIVE FB RECSIZE(40) BLKSIZE(6200));
27.      DCL DTREND OUTPUT FILE RECORD SEQL BUF ENV(TOTAL
28.          CONSECUTIVE FB RECSIZE(12) BLKSIZE(6228));
29.      DCL (CONTGP,TEMP_HERD,TEMP_DAM,SIREID,
29.5       TEMP_DAMNO,TEMP_DAMYR INIT(0)) FIXED BIN(31,0);
30.      DCL (TOTAL_SS,DAM_SS,CG_SS,BYR_SS,SIRE_SS) INIT(0) FLOAT DEC(16);
31.      DCL (Z1PY INIT(0),DAMSOL INIT(0),ERROR_VAR INIT(0)) FLOAT DEC(16);
32.      DCL ALPHA INIT(7) FLOAT DEC;
33.      DCL (I,J,K,L) INIT(0) FIXED BIN;
34.      DCL (#DAMS,#SIRES,#BYR,#CGS,#TCGS,#RECS) INIT(0) FIXED BIN;
35.      DCL (CHAR,DEC,SUBSTR) BUILTIN;
36.      DCL EOFSW BIT(1) INIT('1'B);
37.      DCL EOFSW2 BIT(1) INIT('1'B);
38.      DCL EOFSW3 BIT(1) INIT('1'B);
39.      DCL EOFSW4 BIT(1) INIT('1'B);
40.      DCL (N,P,Q,R,S,T,U) PTR;
41.
42.      /* For genetic groups.  */
43.
44.      DCL 1 BYR_SOL BASED(N),
45.         (2 BREED,
46.          2 EQN#,
47.          2 PAD) FIXED BIN(31,0),
48.         (2 SOLUTION,
49.          2 PAD2) FLOAT DEC(16);
50.
51.      DCL 1 SIRE_SOL BASED(P),
52.         (2 SIRE_REG,
53.          2 BYR,
54.          2 EQN#) FIXED BIN(31,0),
55.         (2 R_SQ,
56.          2 EPN,
57.          2 SHAT,
58.          2 UHAT,
```

```
59.           2 R_SQ2) FLOAT DEC(16);
60.
61.       DCL 1 CG_SOL BASED(Q),
62.          (2 PAD,
63.           2 HERD,
64.           2 CG) FIXED BIN(31,0),
65.           2 SOLUTION FLOAT DEC(16);
66.
67.       DCL 1 REC_OUT BASED(R),
68.          (2 DAM_YR,
69.           2 HERD,
70.           2 DAM_REG) FIXED BIN(31,0),
71.          (2 SOLUTION,
72.           2 DIAG) FLOAT DEC(16);
73.
74.       DCL 1 DATA_IN BASED(S),
75.           2 PAD CHAR(2),
76.           2 HERD CHAR(8),
77.           2 SEX PIC'9',
78.           2 WEAN_YR PIC'99',
79.           2 WEAN_DAY PIC'999',
80.           2 WEAN_MGT PIC'9',
81.           2 YEAR_DATE PIC'(5)9',
82.           2 YEAR_TYPE PIC'9',
83.           2 YEAR_MGT PIC'9',
84.           2 DATA_SOURCE PIC'9',
85.           2 SIRE_REG PIC'(8)9',
86.           2 SIRE_YRGP PIC'(2)9',
87.           2 PROGENY_REC PIC'(4)9',
88.           2 DAM_REG PIC'(8)9',
89.           2 DAM_YR PIC'99',
90.           2 DAM_DAY PIC'(3)9',
91.           2 CALF_REG PIC'(8)9',
92.           2 DAM_NO PIC'(2)9';
93.
94.       DCL 1 TRD_OUT BASED(T),
95.          (2 HERD,
96.           2 YR,
97.           2 CG_ID,
98.           2 DAM_ID,
99.           2 SIRE_ID,
100.          2 FREQ) FIXED BIN(31,0),
101.         (2 PH_SOL,
102.          2 CG_SOL,
103.          2 DAM_SOL,
104.          2 SIRE_SOL) FLOAT DEC;
104.1
105.      DCL 1 DTRD_OUT BASED(U),
106.         (2 HERD,
107.          2 ID) FIXED BIN(31,0),
108.          2 SOL FLOAT DEC;
109.
110.
111.      DCL 1 SIRELIST(990),
112.         (2 REG INIT((990)0),
113.          2 BYR INIT((990)0)) FIXED BIN(31,0),
114.          2 SOLUTION INIT((990)0) FLOAT DEC(16);
115.
116.      DCL 1 CGLIST(1400),
117.          2 CG INIT((1400)0) FIXED BIN(31,0),
```

```
118.          2 SOLUTION INIT((1400)0) FLOAT DEC(16);
119.
120.       DCL 1 BYRLIST(20),
121.          2 BYR INIT((20)0) FIXED BIN,
122.          2 SOLUTION INIT((20)0) FLOAT DEC(16);
123.
124.       /*************************************************/
125.       /*  Read in sire solutions.                      */
126.       /*************************************************/
127.
128.       ON ENDFILE(SIREIN) EOFSW2='0'B;
129.       READ FILE(SIREIN) SET(P);
130.
131.       DO WHILE(EOFSW2);
132.         #SIRES=#SIRES+1;
133.         SIRELIST.REG(#SIRES)=SIRE_SOL.SIRE_REG;
134.         IF SIRE_SOL.BYR=0 THEN
135.               SIRELIST.BYR(#SIRES)=64;
136.         ELSE SIRELIST.BYR(#SIRES)=63+SIRE_SOL.BYR;
137.         SIRELIST.SOLUTION(#SIRES)=SIRE_SOL.SHAT;
138.         READ FILE(SIREIN) SET(P);
139.       END;
140.
141.
142.       /*************************************************/
143.       /*  Read in genetic group solutions.             */
144.       /*************************************************/
145.
146.       ON ENDFILE(BYRIN) EOFSW3='0'B;
147.       READ FILE(BYRIN) SET(N);
148.
149.       DO I=1 TO 18;
150.         #BYR=#BYR+1;
151.         BYRLIST.BYR(I)=63+I;
152.         BYRLIST.SOLUTION(I)=BYR_SOL.SOLUTION;
153.         READ FILE(BYRIN) SET(N);
154.       END;
155.
156.       /*************************************************/
157.       /*  Read in contemporary group solutions.        */
158.       /*************************************************/
159.
160.       ON ENDFILE(CGIN) EOFSW='0'B;
161.       READ FILE(CGIN) SET(Q);
162.
163.       ON ENDFILE(DATAIN) EOFSW4='0'B;
164.       READ FILE(DATAIN) SET(S);
165.
166.       DO WHILE(EOFSW);
167.         #CGS=0;
168.         CGLIST.CG=0;
169.         CGLIST.SOLUTION=0;
170.         TEMP_HERD=CG_SOL.HERD;
171.           DO WHILE(EOFSW & TEMP_HERD=CG_SOL.HERD);
172.             #CGS=#CGS+1;
173.             #TCGS=#TCGS+1;
174.             CGLIST.CG(#CGS)=CG_SOL.CG;
175.             CGLIST.SOLUTION(#CGS)=CG_SOL.SOLUTION;
176.             READ FILE(CGIN) SET(Q);
177.          END;
```

```
178.
179.          TEMP_DAM=DATA_IN.DAM_REG;
179.1         TEMP_DAMYR=DATA_IN.DAM_YR;
179.2         TEMP_DAMNO=DATA_IN.DAM_NO;
180.
181.          DO WHILE(EOFSW4 & TEMP_HERD=DATA_IN.HERD);
182.
183.             #DAMS=#DAMS+1;
184.
185.             DO WHILE(EOFSW4 & TEMP_DAM=DATA_IN.DAM_REG);
186.
187.               #RECS=#RECS+1;
188.               TOTAL_SS=TOTAL_SS+(DATA_IN.PROGENY_REC)**2;
189.               CONTGP=DATA_IN.SEX||DATA_IN.WEAN_YR||DATA_IN.WEAN_DAY||
190.                 DATA_IN.WEAN_MGT||DATA_IN.DATA_SOURCE;
191.               I=FIND(CONTGP,1,#CGS,1);
192.               CG_SS=CG_SS+CGLIST.SOLUTION(I)*DATA_IN.PROGENY_REC;
193.               SIREID=DATA_IN.SIRE_REG;
194.               J=FIND(SIREID,1,#SIRES,2);
195.               K=FIND(SIRELIST.BYR(J),1,#BYR,3);
196.               BYR_SS=BYR_SS+BYRLIST.SOLUTION(K)*DATA_IN.PROGENY_REC;
197.               SIRE_SS=SIRE_SS+SIRELIST.SOLUTION(J)*
198.                     DATA_IN.PROGENY_REC;
199.
200.       /*************************************************/
201.       /*  Determine dam solutions and dam_ss.          */
202.       /*                                               */
203.       /*************************************************/
204.
205.               Z1PY=Z1PY+DATA_IN.PROGENY_REC;
206.               DAMSOL=DAMSOL+(1/(DATA_IN.DAM_NO+ALPHA))*(DATA_IN.PROGENY_REC
207.                -CGLIST.SOLUTION(I)-BYRLIST.SOLUTION(K)-SIRELIST.SOLUTION(J));
208.
209.       /*************************************************/
210.       /*  Output for within herd by year data for      */
211.       /*  phenotypic, genetic, and environmental       */
212.       /*  trend lines.                                 */
213.       /*************************************************/
214.
215.               LOCATE TRD_OUT FILE(TREND) SET(T);
216.                 TRD_OUT.HERD=DATA_IN.HERD;
217.                 TRD_OUT.YR=DATA_IN.WEAN_YR;
218.                 TRD_OUT.CG_ID=CONTGP;
219.                 TRD_OUT.PH_SOL=DATA_IN.PROGENY_REC;
220.                 TRD_OUT.DAM_ID=TEMP_DAM;
221.                 TRD_OUT.SIRE_ID=DATA_IN.SIRE_REG;
222.                 TRD_OUT.FREQ=1;
223.                 TRD_OUT.CG_SOL=CGLIST.SOLUTION(I);
224.                 TRD_OUT.DAM_SOL=0;
225.                 TRD_OUT.SIRE_SOL=SIRELIST.SOLUTION(J)
225.5                      +BYRLIST.SOLUTION(K);
226.
227.               READ FILE(DATAIN) SET(S);
228.             END;
229.
230.              DAM_SS=DAM_SS+DAMSOL*Z1PY;
231.              Z1PY=0;
232.
233.             LOCATE  REC_OUT FILE(RECOUT) SET(R);
234.               REC_OUT.DAM_YR=TEMP_DAMYR;
```

```
235.              REC_OUT.HERD=TEMP_HERD;
236.              REC_OUT.DAM_REG=TEMP_DAM;
237.              REC_OUT.SOLUTION=DAMSOL;
238.              REC_OUT.DIAG=TEMP_DAMNO+ALPHA;
239.
240.
241.              LOCATE DTRD_OUT FILE(DTREND) SET(U);
242.                DTRD_OUT.HERD=TEMP_HERD;
243.                DTRD_OUT.ID=TEMP_DAM;
244.                DTRD_OUT.SOL=DAMSOL;
245.
247.              TEMP_DAM=DATA_IN.DAM_REG;
247.2             TEMP_DAMYR=DATA_IN.DAM_YR;
247.3             TEMP_DAMNO=DATA_IN.DAM_NO;
248.              DAMSOL=0;
249.           END;
250.        END;
251.
252.        ERROR_VAR=(TOTAL_SS-CG_SS-DAM_SS-BYR_SS-SIRE_SS)/
253.                  (#RECS-(#CGS+#BYR-1));
254.        PUT LIST('TOTAL SS =                          ',TOTAL_SS) SKIP;
255.        PUT LIST('CONTEMPORARY GROUP SS =              ',CG_SS) SKIP;
256.        PUT LIST('DAM SS =                             ',DAM_SS) SKIP;
257.        PUT LIST('GENETIC YEAR GROUP SS =                 ',BYR_SS) SKIP;
258.        PUT LIST('SIRE SS =                           ',SIRE_SS) SKIP;
259.        PUT LIST('ERROR VARIANCE =                       ',ERROR_VAR) SKIP;
260.        PUT LIST('NUMBER OF CONTEMPORARY GROUPS = ',#TCGS) SKIP;
261.        PUT LIST('NUMBER OF SIRES =               ',#SIRES) SKIP;
262.        PUT LIST('NUMBER OF GENETIC GROUPS =      ',#BYR) SKIP;
263.        PUT LIST('TOTAL NUMBER OF RECORDS =       ',#RECS) SKIP;
264.        PUT LIST('TOTAL NUMBER OF DAMS =          ',#DAMS) SKIP;
265.
266.        FIND: PROC(ID,START,STOP,SW) RETURNS(FIXED BIN) REORDER;
267.          DCL (START,STOP,SW,HIGH,LOW,MID,MATCH INIT(0)) FIXED BIN;
268.          DCL ID FIXED BIN(31,0);
269.          LOW=START;
270.          HIGH=STOP;
271.
272.          IF SW=1 THEN DO;
273.             DO WHILE(HIGH>=LOW & MATCH=0);
274.                MID=(LOW+HIGH)/2;
275.                SELECT;
276.                  WHEN (ID=CGLIST.CG(MID)) MATCH=MID;
277.                  WHEN (ID>CGLIST.CG(MID)) LOW=MID+1;
278.                  WHEN (ID<CGLIST.CG(MID)) HIGH=MID-1;
279.                END;
280.             END;
281.           END;
282.
283.          IF SW=2 THEN DO;
284.             DO WHILE(HIGH>=LOW & MATCH=0);
285.                MID=(LOW+HIGH)/2;
286.                SELECT;
287.                  WHEN (ID=SIRELIST.REG(MID)) MATCH=MID;
288.                  WHEN (ID>SIRELIST.REG(MID)) LOW=MID+1;
289.                  WHEN (ID<SIRELIST.REG(MID)) HIGH=MID-1;
290.                END;
291.             END;
292.           END;
293.
```

```
294.          IF SW=3 THEN DO;
295.             DO WHILE(HIGH>=LOW & MATCH=0);
296.                MID=(LOW+HIGH)/2;
297.                SELECT;
298.                  WHEN (ID=BYRLIST.BYR(MID)) MATCH=MID;
299.                  WHEN (ID>BYRLIST.BYR(MID)) LOW=MID+1;
300.                  WHEN (ID<BYRLIST.BYR(MID)) HIGH=MID-1;
301.                END;
302.             END;
303.           END;
304.         RETURN (MATCH);
305.        END FIND;
306.        END SS_SQRS;
307.       /*
308.       //GO.BYRIN DD DSN=A.U3274.WU12,UNIT=3330,DISP=(OLD,KEEP),
309.       // VOL=SER=AGS208
310.       //GO.SIREIN DD DSN=A.U3274.WACCLIST,UNIT=3330,DISP=(OLD,KEEP),
311.       // VOL=SER=AGS208
312.       //GO.CGIN DD DSN=A.U3274.CONTSOL,UNIT=3330,DISP=(OLD,KEEP),
313.       // VOL=SER=AGS208
314.       //GO.DATAIN DD DSN=A.U3274.WREC2,UNIT=3330,DISP=(OLD,KEEP),
315.       // VOL=SER=AGS208
316.       //GO.RECOUT DD DSN=A.U3274.DAMSOL,UNIT=3330,DISP=(NEW,CATLG),
317.       // SPACE=(TRK,(50,20),RLSE),VOL=SER=AGS208,
318.       // DCB=(RECFM=FB,LRECL=28,BLKSIZE=6216)
319.       //GO.TREND DD DSN=&&TEMP1,UNIT=SCRTCH,DISP=(NEW,PASS),
320.       // SPACE=(TRK,(50,20),RLSE),DCB=(RECFM=FB,LRECL=40,BLKSIZE=6200)
321.       //GO.DTREND DD DSN=&&TEMP2,UNIT=SCRTCH,DISP=(NEW,PASS),
322.       // SPACE=(TRK,(30,10),RLSE),DCB=(RECFM=FB,LRECL=12,BLKSIZE=6228)
323.       //S3  EXEC  MATCHUP
324.       //MASTERIN DD DSN=&&TEMP1,DISP=(OLD,DELETE),UNIT=SCRTCH
325.       //SELECT  DD  DSN=&&TEMP2,DISP=(OLD,DELETE),UNIT=SCRTCH
326.       //MATCH  DD  DSN=A.U3274.PHTRD,DISP=(NEW,CATLG),UNIT=3330,
327.       //     VOL=SER=AGS208,SPACE=(TRK,(50,20),RLSE),
328.       //     DCB=(RECFM=FB,LRECL=40,BLKSIZE=6200)
329.       //SYSIN  DD  *
330.       MASK(1,1,4/13,5,4//33,9,4),ALL
331.       //ST4 EXEC SYMSORT,TIME=1,TRACKS=100
332.       //SORTIN DD DSN=A.U3274.PHTRD,UNIT=3330,DISP=(OLD,KEEP),
333.       // VOL=SER=AGS208
334.       //SORTOUT DD DSN=A.U3274.CGTRD,UNIT=3330,VOL=SER=AGS208,
335.       // DISP=(NEW,CATLG),SPACE=(TRK,(50,20),RLSE),
336.       // DCB=(RECFM=FB,LRECL=40,BLKSIZE=6200)
337.       //SYSIN DD *
338.         SORT FIELDS=(1,12,BI,A)
339.         SUM FIELDS=(21,4,BI)
340.       /*
341.       //ST5 EXEC SYMSORT,TIME=1,TRACKS=100
342.       //SORTIN DD DSN=A.U3274.PHTRD,UNIT=3330,DISP=(OLD,KEEP),
343.       // VOL=SER=AGS208
344.       //SORTOUT DD DSN=A.U3274.DAMTRD,UNIT=3330,VOL=SER=AGS208,
345.       // DISP=(NEW,CATLG),SPACE=(TRK,(50,20),RLSE),
346.       // DCB=(RECFM=FB,LRECL=40,BLKSIZE=6200)
347.       //SYSIN DD *
348.         SORT FIELDS=(1,8,BI,A,13,4,BI,A)
349.         SUM FIELDS=(21,4,BI)
350.       /*
351.       //ST6 EXEC SYMSORT,TIME=1,TRACKS=100
352.       //SORTIN DD DSN=A.U3274.PHTRD,UNIT=3330,DISP=(OLD,KEEP),
353.       // VOL=SER=AGS208
```

```
354.     //SORTOUT DD DSN=A.U3274.SIRETRD,UNIT=3330,VOL=SER=AGS208,
355.     // DISP=(NEW,CATLG),SPACE=(TRK,(50,20),RLSE),
356.     // DCB=(RECFM=FB,LRECL=40,BLKSIZE=6200)
357.     //SYSIN DD *
358.       SORT FIELDS=(1,8,BI,A,17,4,BI,A)
359.       SUM FIELDS=(21,4,BI)
360.     /*
361.     //ST7 EXEC SYMSORT,TIME=1,TRACKS=100
362.     //SORTIN DD DSN=A.U3274.PHTRD,UNIT=3330,DISP=(OLD,KEEP),
363.     // VOL=SER=AGS208
364.     //SORTOUT DD DSN=A.U3274.PHTRD2,UNIT=3330,VOL=SER=AGS208,
365.     // DISP=(NEW,CATLG),SPACE=(TRK,(50,20),RLSE),
366.     // DCB=(RECFM=FB,LRECL=40,BLKSIZE=6200)
367.     //SYSIN DD *
368.       SORT FIELDS=(1,8,BI,A)
370.     /*
```